VIRAL GENOMICS FOR IDENTIFICATION OF SIGNALS OF DISEASE IN ENVIRONMENTAL SAMPLES

By

Camille McCall

A DISSERTATION

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

Environmental Engineering—Doctor of Philosophy

2020

ABSTRACT

VIRAL GENOMICS FOR IDENTIFICATION OF SIGNALS OF DISEASE IN ENVIRONMENTAL SAMPLES

By

Camille McCall

Viruses have been responsible for some of the most notorious outbreaks and pandemics in modern history. With increases in urbanization and global transportation we can expect viruses to remain a major concern both now and in the future. It is important to establish new ways to monitor virus circulation in communities and forecast the onset of a potential outbreak. Since centralized wastewater treatment facilities have the capacity to collect wastewater from thousands or millions of inhabitants per day, a wastewater-based epidemiology (WBE) system can serve as an indicator of population health. This work aims to identify signals of disease in wastewater and potential for early detection of viral disease outbreaks in communities using molecular approaches and optimized sequencing strategies.

Untreated wastewater samples were collected from a wastewater treatment plant situated in a large metropolitan area in the United States. Viral pathogens were identified in samples using qPCR and viral metagenomics (viromics). Mechanistic modeling and statistical approaches were used to determine the potential for early detection of select viral diseases. Public health data was applied to confirm the incidence of diseases associated with pathogens found in wastewater. Overall findings from this work suggests that WBE can be used to detect early peaks in select viral disease cases within a community before health care facilities are notified. Optimized metagenomic approaches and qPCR suggest that important viruses classified as enteric, respiratory, bloodborne, vector-borne and others are excreted in wastewater and can be monitored to make inferences about

population health and potential for emerging disease outbreaks. Moreover, results indicate that specific public reporting of important viruses causing flu-like and gastrointestinal illness can enhance the efficacy of WBE to assess the burden of pathogens causing nonspecific illnesses. WBE along with molecular approaches and viral metagenomics has the potential to revolutionize public health and government responses to outbreaks. New approaches of this nature can be implemented in communities across the globe in an effort to mitigate the impacts of viral disease outbreaks on the economy and public health. I would like to dedicate this dissertation to my nieces Ava and Malia and my goddaughter Tamia who have inspired me with their unconfined trust and belief in me. It is my hope that my efforts will encourage them to accomplish all that they desire.

ACKNOWLEDGEMENTS

A most sincere thanks goes to my advisor Dr. Irene Xagoraraki for her constant guidance, support, and investment in my ideas, work, and development. Thank you to my committee members, Dr. Susan Masten, Dr. Alison Cupples, and Dr. Shin-Han Shiu for their support, availability, and thoughtful feedback. Many thanks to my lab mate Evan O'Brien for his support and willingness to assist with the work presented in this dissertation. I very much appreciate my lab mates Huiyun Wu and Brijen Miyani for their constant support, friendship, and collaborations. Thank you to my previous supervisors AnnMarie Schneider and Dr. Jenahvive Morgan for their support, invaluable advice, and encouragement.

I am most grateful for all my friends and family who have been there to uplift me throughout my program. My heart is filled with gratitude for my sisters and confidants Jessica McCall- Kailimai, Dominique Cunningham, Lina Smith, Tula Ngasala, and Joyce Mujunga. I am blessed to have them. A special thank you to my mother Donella Abram who has not withheld anything and has taught me that all things are possible. I am grateful for my beloved dog Sam. She has been with me throughout all my academic years never leaving my side.

I am forever indebted to God who has planted my feet firmly on the ground, leveled the mountains, and lifted the valleys to bring this moment into existence. This journey has been like a river, sometimes rough, sometimes calm, but always moving. The abundance of love and support from everyone is what keeps me moving.

LIST OF TABLES	ix
LIST OF FIGURES	x
INTRODUCTION	1
REFERENCES	6
CHADTED 1. METAGENOMIC ADDOACHES FOD DETECTING VIDAL DIVEDSITY I	N
WATER ENVIRONMENTS	11 11
Abstract	. 11
1. Introduction	. 12
2. Viral Pathogen Discovery and Diversity of Viruses in Water Environments Using	
Metagenomics	. 15
3. Methodology in Viral Metagenomics	. 17
4. Virus Concentration and Enrichment	. 20
5. Quality Analysis of Viral-Related Sequence Reads	. 24
6. Assembly of Viromes	. 25
7. Alignment and Taxonomic Classification of Viromes	. 27
8. Limitations of Next-Generation Sequencing for Viral Diversity and Pathogen Detection	. 29
9. Summary and Conclusions	. 31
REFERENCES	. 32
CHAPTER 2: EARLY DETECTION OF A HEPATITIS OUTBREAK IN AN URBAN	
COMMUNITY USING WASTEWATER-BASED EPIDEMIOLOGY	. 39
Abstract	. 39
1. Introduction	. 40
2. Methods	. 42
2.1. Study Area and Wastewater Sample Collection	. 42
2.2. Sample Processing and Virus Isolation	. 45
2.3. Preparation of HAV Standards	. 45
2.4. Quantitative RT-PCR and Limit of Detection	. 46
2.5. Next Generation Sequencing and Metagenomic Analysis	. 47
2.6. Creatinine Analysis	. 49
2.7. Precipitation Data Collection	. 49
2.8. Clinical Data Collection	. 50
2.9. Mechanistic Modeling for Selection of Hepatitis A Cases	. 51
2.10. Statistical Analyses	. 53
3. Results	. 56
3.1. Environmental Surveillance of HAV and Correlation to Clinical Cases	. 56
3.2. Effects of Population, Precipitation, and Disease on HAV Concentrations in	
Wastewater Samples	. 59
3.3. Metagenomic Analysis of Viral Hepatitis	. 60
4. Discussion	. 61

TABLE OF CONTENTS

4.1. Environmental Surveillance of HAV and Connection to Clinical Data	61
4.2. Factors Influencing HAV Concentrations in Wastewater	63
4.3. Metagenomic Analysis of Viral Hepatitis in Wastewater	66
4.4. WBE for Early Detection of Viral Disease Outbreaks	68
5. Conclusion	69
Acknowledgements	
APPENDICES	
APPENDIX A2: Supplementary Methods	
APPENDIX B2: Supplementary Table and Figures	
REFERENCES	
CHAPTER 3: IDENTIFICATION OF MULTIPLE POTENTIAL VIRAL DISEASES IN	А
LARGE URBAN CENTER USING WASTEWATER SURVEILLANCE	99
Abstract	
1. Introduction	100
2. Methods	102
2.1 Study Area and Wastewater Sample Collection	102
2.1. Study Thea and Wastewater Sample Concerton	103
2.2. Sumple Processing and Virus Isolation	104
2.3. Nettagenomic Analysis	104
2.3.1. Sampling Processing and Kandom amplification	104
2.3.2. Next Generation Sequence Frocessing	104
2.5.5. Sequence Analysis and Taxonomic Annotation	104
2.4. Quantification of Steer level Correct	107
2.4.1. Preparation of Standard Curves	107
2.4.2. qPCK and $K1$ -qPCK	108
2.5. Health Data Collection	109
2.6. Statistical and Cluster Analysis	110
3. Results	110
3.1. Classification of Human Viral Pathogens in Wastewater	111
3.2. Comparison of Human Viruses and Clinical Data	119
3.3. Quantitative Screening for Select Human Viral Pathogens	119
4. Discussion	121
4.1. Classification of Human Viruses in Wastewater and Clinical Data Comparison	121
4.2. Screening of Select Human Viral Pathogens	127
5. Conclusion	128
Acknowledgements	129
APPENDIX	130
REFERENCES	137
CHAPTER 4: ASSESSMENT OF CALICIVIRUSES AND OTHER ENTERIC VIRUSE	S IN A
LARGE METROPOLITAN AREA USING WASTEWATER SURVEILLANCE	145
Abstract	145
1. Introduction	146
2. Methods	147
2.1 Study Area and Wastewater Sample Collection	147
2.2 Sample Processing and Virus Isolation	148
=-=	10

2.3. Metagenomic Analysis	149
2.3.1. Random Amplification and Next Generation Sequence Processing	149
2.3.2. Sequence Analysis and Taxonomic Anotation	150
2.4. Quantitative PCR	151
2.5. Health Data Collection	152
2.6. Statistical Analysis	152
3. Results	153
3.1. Calicivirus detection in Wastewater	153
3.2. Metagenomic Screening of NoV and SaV	155
4. Discussion	156
4.1. Calicivirus Quantification in Wastewater and Clinical Presence	156
4.2. Genogroup Classification of NoV and SaV and Diversity of Enteric Viruses in	
Wastewater	159
5. Conclusion	161
Acknowledgements	162
REFERENCES	163
CONCLUSIONS AND SIGNIFICANCE	170

LIST OF TABLES

Table 2. 1. Average measured and collected environmental and disease data during for each sampling date during study period. 58
Table B2. 1. Summary of sampling schedule and locations. 74
Table B2. 2. Wastewater physio-chemical characteristics per sampling date and location for sampling year 1. 76
Table B2. 3. Wastewater physio-chemical characteristics per sampling date and location for sampling year 2. 77
Table 3. 1. Total number of contigs per virus family with associated host for 18 sequenced samples.
Table 3. 2. Summary of human viral pathogens detected in wastewater and their associated diseasereported in the Michigan Disease Surveillance System (MDSS)Weekly Surveillance Reports(WSR). Associated disease is considered if at least one case was reported during the sampling year(2017-2018)
Table A3. 1. Primers and probes for select viruses
Table A3. 2. Table A3.2. Summary of reads produced from sequenced cDNA samples and metagenomic alignment statistics. Viruses assigned to human viral group include Riboviruses and virus assigned to the root. 132
Table A3. 3. Reportable viral associated diseases in Michigan. Number of disease cases forMichigan, Detroit City, Wayne County, Oakland County, and Macomb County in 2017 and 2018.umber represent probable and confirmed cases.133
Table 4. 1. Average concentrations of NoV GII and SaV per sampling date along with number ofGI and noroviruses cases reported within the service community.154

LIST OF FIGURES

 Figure B2. 9. Distribution of hepatitis A virus concentrations per sampling year for each sampling event. Time is indicated as sampling week. Lines produced from linear regression analysis. Shaded regions represent 95% confidence interval.
 86

 Figure B2. 10. Comparison of Hepatitis A concentrations and number of cases for sampling year 1 and 2. Line representation linear regression lines.
 87

 Figure B2. 11. Summary of stepwise multiple linear regression analysis. Final model. Regression analysis performed in R.
 88

 Figure B2. 12. Diagnostics plots obtained from the final linear regression model.
 89

 Figure B2. 13. Assessment of remaining outliers using Cook's distance (A) and Cook's distance vs. leverage (B) plots. All outliers were considered quality measured concentrations.
 90

 Figure B2. 14. Total number of contigs assigned to each virus group during metagenomic analysis. Total number includes all 18 samples for sampling year 1. Further analysis was performed on the Riboviria realm.
 91

Figure 3. 1. Metagenomic workflow for human virus identification in wastewater samples..... 106

Figure 3. 2. ssRNA (a) and DNA	(b) virus diversity and relative abundance in wastewater samples.

Figure 3. 4. (a) Heatmap of human genus virus diversity and normalized abundance in each sample. White cells indicate absence of associated virus in related sample. Virus genera are in descending order according to abundance. Heatmap was produced in R. (b) Proportion of human virus types Figure A3. 1. Tukey's post hoc statistical test results. Plot displays the comparison of mean virus concentrations. Intervals that lie on the dotted line mean that there is no difference between the mean concentrations between those two viruses. Intervals that lie to the left of the dotted line means that the concentration of the first virus listed in the pair is significantly less than the second virus. If interval is on the right side of the dotted line the first virus is significantly greater than the Figure 4. 1. Boxplots of norovirus genogroup II (NoV GII) and sapovirus (SaV) average concentrations in wastewater samples per sampling date during years one (A) and two (B). Median Figure 4. 2. Normalized abundance of norovirus and sapovirus in metagenomic samples...... 155 Figure 4. 3. Proportion of enteric viruses detected in wastewater samples collected during sampling

INTRODUCTION

Viral disease outbreaks have substantial economic impacts on communities and healthcare facilities. A single outbreak can cost the U.S millions of dollars in diagnostics tests, treatment, hospitalizations, and losses in productivity. Developing communities are especially at risk since the high financial stress of outbreaks can cause collapses in community and healthcare infrastructures. Early detection strategies and systems to identify warning signs of viral disease outbreaks are rapidly evolving in an effort to prevent the next epidemic or even pandemic.

Several studies have explored various methodologies in early disease outbreak detection including evaluating absenteeism of k-12 children (Besculides et al., 2005), the surveillance of dead bird clusters (Mostashari et al., 2003), the number of over-the-counter drug sales (Das et al., 2005) and search engine query data (Althouse et al., 2011; Ginsberg et al., 2009; Pervaiz et al., 2012; Verma et al., 2018). However, these approaches do not provide direct measurements of viral presence and concentration patterns or explore the potential for novel virus outbreaks. Given that viruses do not replicate outside of a human host and can remain stable in water for significant periods of time, sewage systems are gaining popularity in outbreak investigations and surveillance of important pathogens.

Untreated (raw) wastewater harbors a wealth of information about the community in the sewage catchment area. Wastewater has been used to determine the exposure of physical, biological and chemical stressors within communities. Such sewage-based application was first proposed in 2001 (Daughton, 2001) and used to evaluate illicit drug use in 2005 (Zuccato et al., 2005). This approach, termed wastewater-based epidemiology (WBE), has since been used to evaluate the

health status of large communities. In WBE, raw sewage is considered analogous to a clinical urine or stool sample from a large population. This representative sample can be used to assess or even predict diseases and community health trajectories by directly measuring concentrations for biomarkers of concern. Several studies have probed sewage for answers to health effects of temperature on communities (Phung et al., 2017), phthalate (González-Mariño et al., 2017) and heavy metal (Markosian and Mirzoyan, 2019) exposure, community diet trends (Venkatesan et al., 2019), the incidence of foodborne bacterial illness (Yan et al., 2018), and virus circulation in communities (Bisseux et al., 2018; Brouwer et al., 2018; Kamel et al., 2010; Kokkinos et al., 2011; La Rosa et al., 2014).

Provided the nature of enteric viruses, they are considered model viruses for WBE. These viruses are typically ones that infect the intestinal tract and can be transmitted via the fecal-to-oral route and enter waste streams through human feces and urine. Enteric viruses can disseminate rapidly in dense populations resulting in large-scale epidemics. To name a few, hepatitis A and E viruses, norovirus, sapovirus, enterovirus, astroviruses, adenovirus, and rotavirus are commonly monitored in sewage (Bisseux et al., 2018; Bonanno Ferraro et al., 2020; Brouwer et al., 2018; Kamel et al., 2010; Kokkinos et al., 2011).

Limited studies have investigated WBE for the surveillance of non-enteric viruses. Influenza, a notorious seasonal respiratory pathogen, has been discovered in human feces (Arena et al., 2012; Hirose et al., 2016), highlighting the potential for its appearance in wastewater. Heijnen and Medema, 2011 observed one positive influenza A sample collected from raw wastewater during the 2009 H1N1 pandemic although the virus isolated did not appear to be related to the pandemic

type. Additionally, WBE is at the forefront of the coronavirus disease 2019 (COVID-19) pandemic as researchers work on exploiting the sewage with new and conventional technologies to predict the second wave of infections (Ahmed et al., 2020; Mao et al., 2020; Medema et al., 2020; Orive et al., 2020; Wu et al., 2020; Wurtzer et al., 2020).

Conventional approaches to sewage exploration include polymerase chain reaction (PCR) techniques and amplicon sequencing approaches. Although much work is still required, pairing these conventional approaches with enriched virus concentration methods (e.g. electropositive filtration, ultrafiltration, flocculation, and nuclease treatment) (Hall et al., 2014; McCall and Xagoraraki, 2019) and public health records has delivered evidence that WBE can provide insight into human microbiomes. The invention of next generation sequencing (NGS) technologies coupled with metagenomics makes it possible to survey a range of viruses in sewage and provide insight into viral threats that warrant further attention. NGS is the sequencing of genomes within a sample, whether environmental or human, that can be further analyzed using computational workflows and tools, genomic databases, and pipelines (metagenomics). Metagenomic analysis focused solely on viruses is also termed viral metagenomics or viromics (Roux et al., 2017).

A recent study using viromics evaluated viral diversity in toilet waste from 19 different international flights. The diversity of pathogenic viruses followed certain geographical trends between countries and cities concluding that wastewater from airplane toilets can serve as a central hub for global viral disease surveillance (Hjelmsø et al., 2019). Furthermore, metagenomics was implemented to investigate the viral diversity from raw sewage in Kampala, Uganda and the impact of the sewage on receiving waterbodies (O'Brien et al., 2017). Ng et al., 2012 examined

untreated sewage using metagenomics to monitor enteric viruses from four different counties. The authors characterized several novel picorna-like viruses and plant viruses belonging to the *Geminiviridae* family. Likewise, other studies have used viromics to explore virus diversity in wastewater (Martínez-Puchol et al., 2020; McCall and Xagoraraki, 2019). This emphasizes untreated sewage as a focal point for surveying and classifying known and unknown viruses. Although much progress has been made in the arena of viromics for human pathogen surveillance in environmental systems, few or no studies go much beyond diversity and explore relevance to public health data. Chapter three of this work contributes significantly to this field by mapping findings from metagenomic analysis to the incidence of associated diseases in the study area (McCall et al., 2020).

Although fundamental to developments in public and environmental health fields, NGS is not without limitations. Less abundant or small viral genomes can get overlooked by NGS technologies and computational tools due to the presence of more dominant and larger viruses such as bacteriophages – viruses that infect bacteria (McCall and Xagoraraki, 2019). Additionally, viruses can undergo rapid genetic changes, which makes it hard for reference databases to recognize a possibly known but mutated virus. These issues can make NGS generally less sensitive than PCR techniques (Bibby and Peccia, 2013). Therefore, the incorporation of PCR methods is increasingly becoming a gold standard to corroborate NGS studies. Optimization of sampling processing strategies and metagenomic workflows are also gaining attention as researchers look for improved ways to uncover pathogens of concern in environmental reservoirs (Bibby et al., 2011; Martínez-Puchol et al., 2020).

To add to this critical body of work, this study seeks to employ WBE using meteganomic, quantitative PCR (qPCR), and reverse transcription qPCR (RT-qPCR) approaches to forecast viral disease outbreaks and identify human viruses circulating among inhabitants in an urban community. This work is composed of four chapters where Chapter 1 details optimized approaches to detecting human virus diversity in water reservoirs. Chapter 2 employs WBE for early detection of hepatitis A outbreaks and discovery of viral hepatitis types in community sewage using RT-qPCR and metagenomics. Chapters 3 examines untreated wastewater using metagenomics, qPCR, and RT-qPCR to survey relevant human viral pathogens circulating in a large community and their association to reported diseases. Finally, Chapter 4 estimates the burden of sapovirus (SaV) and norovirus (NoV) GII in a large metropolitan area using RT-qPCR. Furthermore, metagenomics was applied to identify NoV and SaV genogroups in wastewater along with other enteric viruses commonly causing acute gastroenteritis. Findings from the abovementioned chapters were used to understand the efficacy of WBE to provide a snapshot of population health.

REFERENCES

REFERENCES

- Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J.W., Choi, P.M., Kitajima, M., Simpson, S.L., Li, J., Tscharke, B., Verhagen, R., Smith, W.J.M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K. V., Mueller, J.F., 2020. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. Sci. Total Environ.
- Althouse, B.M., Ng, Y.Y., Cummings, D.A.T., 2011. Prediction of dengue incidence using search query surveillance. PLoS Negl. Trop. Dis. 5, 1–7.
- Arena, C., Amoros, J.P., Vaillant, V., Balay, K., Chikhi-Brachet, R., Varesi, L., Arrighi, J., Blanchon, T., Carrat, F., Hanslik, T., Falchi, A., 2012. Simultaneous investigation of influenza and enteric viruses in the stools of adult patients consulting in general practice for acute diarrhea. Virol. J. 9.
- Besculides, M., Heffernan, R., Mostashari, F., Weiss, D., 2005. Evaluation of school absenteeism data for early outbreak detection, New York City. BMC Public Health 5, 1–7.
- Bibby, K., Peccia, J., 2013. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. Environ. Sci. Technol.
- Bibby, K., Viau, E., Peccia, J., 2011. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. Lett. Appl. Microbiol.
- Bisseux, M., Colombet, J., Mirand, A., Roque-Afonso, A.M., Abravanel, F., Izopet, J.,
 Archimbaud, C., Peigue-Lafeuille, H., Debroas, D., Bailly, J.L., Henquell, C., 2018.
 Monitoring human enteric viruses in wastewater and relevance to infections encountered in the clinical setting: A one-year experiment in central France, 2014 to 2015.
 Eurosurveillance 23, 1–11.
- Bonanno Ferraro, G., Mancini, P., Veneri, C., Iaconelli, M., Suffredini, E., Brandtner, D., La Rosa, G., 2020. Evidence of Saffold virus circulation in Italy provided through environmental surveillance. Lett. Appl. Microbiol. 70, 102–108.
- Brouwer, A.F., Eisenberg, J.N.S., Pomeroy, C.D., Shulman, L.M., Hindiyeh, M., Manor, Y., Grotto, I., Koopman, J.S., Eisenberg, M.C., 2018. Epidemiology of the silent polio outbreak in Rahat, Israel, based on modeling of environmental surveillance data. Proc. Natl. Acad. Sci. U. S. A. 115, E10625–E10633.
- Das, D., Metzger, K., Heffernan, R., Balter, S., Weiss, D., Mostashari, F., 2005. Monitoring over-the-counter medication sales for early detection of disease outbreaks--New York City. MMWR. Morb. Mortal. Wkly. Rep.

Daughton, C.G., 2001. Illicit drugs in municipal sewage: Proposed new nonintrusive tool to

heighten public awareness of societal use of illicit-abused drugs and their potential for ecological consequences. ACS Symp. Ser.

- Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. Nature 457, 1012–1014.
- González-Mariño, I., Rodil, R., Barrio, I., Cela, R., Quintana, J.B., 2017. Wastewater-Based Epidemiology as a New Tool for Estimating Population Exposure to Phthalate Plasticizers. Environ. Sci. Technol. 51, 3902–3910.
- Hall, R.J., Wang, J., Todd, A.K., Bissielo, A.B., Yen, S., Strydom, H., Moore, N.E., Ren, X., Huang, Q.S., Carter, P.E., Peacey, M., 2014. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. J. Virol. Methods 195, 194–204.
- Heijnen, L., Medema, G., 2011. Surveillance of influenza A and the pandemic influenza A (H1N1) 2009 in sewage and surface water in the Netherlands. J. Water Health 9, 434–442.
- Hirose, R., Daidoji, T., Naito, Y., Watanabe, Y., Arai, Y., Oda, T., Konishi, H., Yamawaki, M., Itoh, Y., Nakaya, T., 2016. Long-term detection of seasonal influenza RNA in faeces and intestine. Clin. Microbiol. Infect. 22, 813.e1-813.e7.
- Hjelmsø, M.H., Mollerup, S., Jensen, R.H., Pietroni, C., Lukjancenko, O., Schultz, A.C., Aarestrup, F.M., Hansen, A.J., 2019. Metagenomic analysis of viruses in toilet waste from long distance flights—A new procedure for global infectious disease surveillance. PLoS One 14, 1–15.
- Kamel, A.H., Ali, M.A., El-Nady, H.G., Aho, S., Pothier, P., Belliot, G., 2010. Evidence of the co-circulation of enteric viruses in sewage and in the population of Greater Cairo. J. Appl. Microbiol.
- Kokkinos, P., Ziros, P., Meri, D., Filippidou, S., Kolla, S., Galanis, A., Vantarakis, A., 2011. Environmental surveillance. An additional/alternative approach for virological surveillance in Greece? Int. J. Environ. Res. Public Health 8, 1914–1922.
- La Rosa, G., Della Libera, S., Iaconelli, M., Ciccaglione, A.R., Bruni, R., Taffon, S., Equestre, M., Alfonsi, V., Rizzo, C., Tosti, M.E., Chironna, M., Romanò, L., Zanetti, A.R., Muscillo, M., 2014. Surveillance of hepatitis A virus in urban sewages and comparison with cases notified in the course of an outbreak, Italy 2013. BMC Infect. Dis. 14, 1–11.
- Mao, K., Zhang, H., Yang, Z., 2020. Can a Paper-Based Device Trace COVID-19 Sources with Wastewater-Based Epidemiology? Environ. Sci. Technol.
- Markosian, C., Mirzoyan, N., 2019. Wastewater-based epidemiology as a novel assessment approach for population-level metal exposure. Sci. Total Environ. 689, 1125–1132.

- Martínez-Puchol, S., Rusiñol, M., Fernández-Cassi, X., Timoneda, N., Itarte, M., Andrés, C., Antón, A., Abril, J.F., Girones, R., Bofill-Mas, S., 2020. Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens. Sci. Total Environ. 713.
- McCall, C., Wu, H., Miyani, B., Xagoraraki, I., 2020. Identification of multiple potential viral diseases in a large urban center using wastewater surveillance. Water Res. 184.
- McCall, C., Xagoraraki, I., 2019. Metagenomic approaches for detecting viral diversity in water environments. J. Environ. Eng. 145.
- Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., Brouwer, A., 2020. Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. Environ. Sci. Technol. Lett. 7, 511–516.
- Mostashari, F., Kulldorff, M., Hartman, J.J., Miller, J.R., Kulasekera, V., 2003. Dead bird clusters as an early warning system for West Nile virus activity. Emerg. Infect. Dis.
- Ng, T.F.F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B.S., Wommack, K.E., Delwart, E., 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J. Virol. 86, 12161–12175.
- O'Brien, E., Nakyazze, J., Wu, H., Kiwanuka, N., Cunningham, W., Kaneene, J.B., Xagoraraki, I., 2017. Viral diversity and abundance in polluted waters in Kampala, Uganda. Water Res.
- Orive, G., Lertxundi, U., Barcelo, D., 2020. Early SARS-CoV-2 outbreak detection by sewagebased epidemiology. Sci. Total Environ.
- Pervaiz, F., Pervaiz, M., Rehman, N.A., Saif, U., 2012. FluBreaks: Early epidemic detection from google flu trends. J. Med. Internet Res. 14, 1–16.
- Phung, D., Mueller, J., Lai, F.Y., O'Brien, J., Dang, N., Morawska, L., Thai, P.K., 2017. Can wastewater-based epidemiology be used to evaluate the health impact of temperature? An exploratory study in an Australian population. Environ. Res. 156, 113–119.
- Roux, S., Emerson, J.B., Eloe-Fadrosh, E.A., Sullivan, M.B., 2017. Benchmarking viromics: An in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ.
- Venkatesan, A.K., Chen, J., Driver, E., Gushgari, A., Halden, R.U., 2019. Assessing the Potential to Monitor Plant-Based Diet Trends in Communities Using a Wastewater-Based Epidemiology Approach. In: ACS Symposium Series.
- Verma, M., Kishore, K., Kumar, M., Sondh, A.R., Aggarwal, G., Kathirvel, S., 2018. Google search trends predicting disease outbreaks: An analysis from India. Healthc. Inform. Res. 24, 300–308.

- Wu, F., Xiao, A., Zhang, J., Moniz, K., Endo, N., Armas, F., Bonneau, R., Brown, M.A., Bushman, M., Chai, P.R., Duvallet, C., Erickson, T.B., Foppe, K., Ghaeli, N., Gu, X., Hanage, W.P., Huang, K.H., Lee, W.L., Matus, M., McElroy, K.A., Nagler, J., Rhode, S.F., Santillana, M., Tucker, J.A., Wuertz, S., Zhao, S., Thompson, J., Alm, E.J., 2020. SARS-CoV-2 titers in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. medRxiv Prepr. Serv. Heal. Sci.
- Wurtzer, S., Marechal, V., Mouchel, J.-M., Maday, Y., Teyssou, R., Richard, E., Almayrac, J.L., Moulin, L., 2020. Evaluation of lockdown impact on SARS-CoV-2 dynamics through viral genome quantification in Paris wastewaters. medRxiv.
- Yan, T., O'Brien, P., Shelton, J.M., Whelen, A.C., Pagaling, E., 2018. Municipal Wastewater as a Microbial Surveillance Platform for Enteric Diseases: A Case Study for Salmonella and Salmonellosis. Environ. Sci. Technol. 52, 4869–4877.
- Zuccato, E., Chiabrando, C., Castiglioni, S., Calamari, D., Bagnati, R., Schiarea, S., Fanelli, R., 2005. Cocaine in surface waters: A new evidence-based tool to monitor community drug abuse. Environ. Heal. A Glob. Access Sci. Source.

CHAPTER 1: METAGENOMIC APPROACHES FOR DETECTING VIRAL DIVERSITY IN WATER ENVIRONMENTS

Published in Journal of Environmental Engineering:

McCall, C., Xagoraraki, I., 2019. Metagenomic approaches for detecting viral diversity in water environments. J. Environ. Eng. 145.

Abstract

Methods for detecting and monitoring known and emerging viral pathogens in the environment are imperative for understanding risk and establishing regulatory standards in environmental and public health sectors. Next-generation sequencing (NGS) has uncovered the diversity of entire microbial populations, enabled discovery of novel organisms, and allowed pathogen surveillance. Metagenomics, the sequencing and analysis of all genetic material in a sample, is a detection method that circumvents the need for cell culturing and prior understanding of microbial assemblies, which are necessary in traditional detection methods. Advancements in NGS technologies have led to subsequent advancements in data analysis methodologies and practices to increase specificity, and accuracy of metagenomic studies. This paper highlights applications of metagenomics in viral pathogen detection, discusses suggested best practices for detecting the diversity of viruses in environmental systems (specifically water environments), and addresses the limitations of virus detection using NGS methods. Information presented in this paper will assist researchers in selecting an appropriate metagenomics approach for obtaining a comprehensive view of viruses in water systems.

1. Introduction

The increase in viral-related diseases, lack of medications to treat viral infections, low infectious dose, and high mortality rates among children, elderly, and the immunocompromised is of global concern. Sewage-treatment plants serve as a significant exposure pathway for viral pathogens. Viral pathogens are known to resist treatment and are released into receiving water bodies, posing a major threat to human health (Wigginton et al., 2015; Xagoraraki et al., 2014). Therefore, detection and surveillance of viruses in water environments is of growing importance. Conventional virus detection methods include indirect methods involving indicator organisms, traditional cell-culture techniques like plaque assays and viral-induced cytopathic effects (CPEs), and molecular techniques such as polymerase chain reaction (PCR). This paper briefly summarizes traditional methods for virus detection and addresses modern techniques, namely, next-generation sequencing, in detail.

Indirect detection methods have been established for suggesting the presence of pathogens in environmental systems. In particular, measuring the concentration of indicator organisms such as coliforms and coliphages is a standard water quality practice for monitoring the presence of enteric pathogens. The survivability of indicator organisms depends on a number of environmental factors, and they are unable to determine the identity of a specific pathogen. Although the concentration of these organisms are suitable for suggesting the occurrence of fecal contamination, they do not always correlate with the presence or absence of viral pathogens. Direct detection methods are required for classifying and monitoring the presence of viral pathogens in environmental samples (Bibby and Peccia, 2013; Ramírez-Castillo et al., 2015).

In the early 1950s, Dulbecco and Vogt, 1954 demonstrated that it was possible to perform plaque assays on human viruses as with bacteriophages, viruses that infect bacteria. Briefly, viruses of varying dilutions are applied to a susceptible cell monolayer. A semisolid media (e.g., agar) is then added to the monolayer to increase viscosity. This restricts the spread of the virus to only neighboring cells within the monolayer. Infected cells lyse and form holes (i.e., plaques) in the monolayer, which can be seen with the naked eye or by staining to increase visibility (Taylor, 2014). Each plaque is known as a plaque-forming unit (PFU) and can be counted and expressed as a concentration. Furthermore, observing CPEs in infected cells is a widely used culture-based method for virus identification in clinical diagnostics. Changes in cell morphology on a cell monolayer are observed microscopically, and the viral agent is identified based on its known CPE. CPEs can appear in various forms, including cell shrinking, swelling, and cell clustering (Leland and Ginocchio, 2007).

Traditional culture-based methods allow for the isolation of purified virus stock and are effective at determining infectivity and concentration. However, these methods are limited by long incubation periods and the inability to culture most viruses. PCR has become a standard molecular method for detecting and quantifying viruses in environmental media. This method bypasses the need for cell culturing, provides rapid detection, and is more efficient for analyzing environmental samples (Xagoraraki et al., 2014). Quantitative PCR (qPCR) enables DNA to be simultaneously quantified and the concentration calculated given a standard curve. Additionally, sequencespecific probes can be used to increase the specificity of DNA amplification. Quantitative reverse transcription (RT-qPCR) involves a collective set of tools to amplify and quantify ribonucleic acid (RNA) viruses such as influenza, rotavirus, and others. In RT-qPCR, RNA is converted into complementary deoxyribonucleic acid (DNA) (cDNA) using an enzyme, reverse transcriptase. The cDNA can then be used as a template for qPCR.

Despite the advances of molecular methods, many of these techniques rely on well-known primers and a priori knowledge of the sample's microbiome. This limits the ability to establish a comprehensive view of viral populations or the discovery of novel pathogens in water and other environmental systems (Fancello et al., 2012). The need for direct human pathogen detection and viral diversity assessments in various ecosystems has led to the development of a new generation of technologies.

Next-generation sequencing (NGS), or high-throughput sequencing (HTS), circumvents the limitations of traditional methods by using advanced technologies to sequence genes or genomes of entire organisms or microbial communities without the need for culturing or prior knowledge of the microbial structure of a sample. Several DNA sequencing methods were introduced in the 1970s (Maxam and Gilbert, 1977; Sanger et al., 1973; Wu and Taylor, 1971). Building on previous methods, in 1977, Frederick Sanger developed a method for DNA sequencing using chain-terminating inhibitors (Sanger et al., 1977). Limitations of the earlier Sanger method included sequencing only up to 1,000 base pairs (bps) of DNA, or short reads (Heather and Chain, 2016). In order to combat this limitation, larger DNA sequences were sheared into smaller fragments, which were then sequenced individually and reassembled through computational methods, a process known as shotgun sequencing (Anderson, 1981; Gardner et al., 1981). Improvements and variations of the Sanger shotgun method have since been introduced into the NGS methods widely used today, thus establishing Sanger as one of the founding pioneers of genomic sequencing.

Whole-genome shotgun (WGS) sequencing is widely used in viral diversity and detection studies. WGS sequencing paired with computational tools and methods (bioinformatics) provide a better understanding of microbial compositions, shifts, and functions. Metagenomics, the sequencing and analysis of all genetic material in a sample (Fancello et al., 2012), has granted unprecedented access into viral communities. Initially, many of the tools used in metagenomics were designed for analyzing bacterial metagenomes. The increase in viral metagenomic studies has led to advancements in established bioinformatics tools as well as new tools and standard strategies for analyzing viral metagenomes (viromes). In this paper, the authors aim to (1) discuss applications of viral metagenomics in terms of viral pathogen discovery and virus diversity in water environments, (2) review standard processes and best practices in viral metagenomic analysis, and (3) discuss limitations of NGS in viral diversity studies.

2. Viral Pathogen Discovery and Diversity of Viruses in Water Environments Using Metagenomics

Metagenomics has proven to be an effective approach in public health and environmental monitoring fields, including disease outbreak analysis, novel pathogen discoveries, and viral diversity studies. Nanopore (Oxford Nanopore Technologies, Oxford, UK) and Illumina (San Diego) sequencing, along with a Nanopore computational pipeline for pathogen detection has been used to identify Ebola, hepatitis C, and chikungunya viruses in human blood samples with high accuracy in as little as 6 h (Greninger et al., 2015). Another recent study used metagenomic deep sequencing on intraocular fluid samples from patients and determined rubella virus to an etiological agent in a patient with bilateral uveitis (Doan et al., 2016). Metagenomic analysis

revealed common viral agents of human respiratory tract infections (i.e., human respiratory syncytial virus, human metapneumovirus, human parainfluenza virus, influenza virus, and human rhinovirus) comprising nearly 90% of viral sequences in nasopharyngeal aspirates samples of infected individuals (Lysholm et al., 2012). Pyrosequencing was used to identity a novel Ebola virus, Bundibugyo ebolavirus, from blood samples of infected patients during an outbreak in western Uganda in 2007. The Bundibugyo ebolavirus was found to be 35%-45% divergent from previously characterized Ebola virus genomes and was therefore difficult to detect using molecular methods. NGS assisted in the identification of this new Ebola virus species with a fatality rate of approximately 36% within 10 days of RNA extraction. The new genome was constructed from the identified novel sequence, and molecular techniques were used to confirm the presence of Bundibugyo ebola virus in isolates of infected patients, thus facilitating rapid isolation and control of the disease (Towner et al., 2008). Moreover, metagenomics has provided a comprehensive look into viral diversity in environmental samples and has identified novel and newly classified viral pathogens in several water and water-related habitats. Bibby and Peccia, 2013 found high occurrences of herpesvirus and papillomavirus, along with newly discovered picornaviruses, in Class B sewage sludge samples using metagenomic analysis. Parechovirus and coronavirus were also detected in biosolids (Bibby et al., 2011), which highlights the risk of biosolids as a means for pathogen transport due to their use in land applications. Metagenomic analysis also identified human viral pathogens rhinovirus, enterovirus, parechovirus, and aichi virus in reclaimed water discharged from public sprinklers, fountains, and spigots at a plant nursery (Rosario et al., 2009). However, these viruses obtained a low percent identity to reference protein sequences, so therefore, careful consideration must be taken when interpreting results. Metagenomics has also been used to assess viral populations in dairy lagoons (Alhamlan et al., 2013), sewage and surface

waters in developed and developing communities (Aw et al., 2014; Fernandez-Cassi et al., 2018; Ng et al., 2012; O'Brien et al., 2017b, 2017a), freshwater lakes (Djikeng et al., 2009), desert ponds (Fancello et al., 2013), coastal (Miranda et al., 2016) and ballast waters (Kim et al., 2015), and in different stages of a conventional wastewater treatment facility (Tamaki et al., 2012). Additionally, a global study on the diversity of viruses in sewage revealed the presence of newly classified klassevirus and a host of novel viruses with at least five different genomes closely related to sequences representing picobirnaviruses (Cantalupo et al., 2011). Table 1 highlights virus concentration techniques, major bioinformatics approaches, and significant findings in viral metagenomic studies in water environments published within the last 10 years. The percentage of human viruses and phages reported from each study represents the percentages of human host viruses and bacteriophages, respectively, in viral-affiliated sequences.

3. Methodology in Viral Metagenomics

A standard workflow containing presequencing and postsequencing processes for viral metagenomic studies in water environments is illustrated in Figure. 1.1. Presequencing practices such as virus concentration and extraction have a significant impact on the diversity of sequenced viral metagenomes. Hall et al., 2014 investigated the impact of commonly used enrichment methods for RNA viruses on the abundance of influenza A and human enterovirus in a synthetic clinical sample. The combination of centrifugation, syringe filtration, and nuclease treatment increased the relative abundance of influenza and enterovirus in the metagenomic dataset by10-fold and 20-fold, respectively. Hjelmsø et al., 2017 evaluated several virus concentration methods and extraction kits as they related to viral metagenomics and found both concentration and

extraction to have a significant impact on viral richness and an interdependent impact on viral specificity.

Standard postsequencing practices for processing raw reads for viral metagenomic analysis include quality filtering, assembling of raw reads into longer contiguous sequences, aligning viromes against reference databases, and taxonomic annotation of aligned sequences. Sequencing platform, assembler, alignment approach, and reference databases are driving factors that influence annotation results. This makes diversity studies difficult and oftentimes incomplete. Implementing best practices during sample processing and analysis can increase the accuracy of virus detection in metagenomic studies.



Figure 1. 1. Suggested WGS workflow for viral metagenomics in water environments.

4. Virus Concentration and Enrichment

Typically, viruses have genomes significantly shorter than some of their host prokaryotic and eukaryotic counterparts. This creates challenges for modern computational tools in the processing of viral sequences because viral signals can be masked by genomes of other organisms. For these reasons, virus concentration is a standard practice in viral metagenomic studies. Ultimately, virus concentration should be performed during sample collection to remove bacteria and other nontargeted material. The virus adsorption-elution (VIRADEL) method is a standard practice for concentrating viruses from environmental waters (U.S. EPA, 2001). This method allows viruses to be captured in large volumes of water and then concentrated in a solution of lesser volume, called eluate. Primary concentration is typically carried out by electropositive or electronegative cartridge filters. These filters make use of the hydrophobic and electrostatic nature of most enteric viruses. Because of the electronegative charge of most viral particles, electronegative filters need to be preconditioned where sampling is performed at a lower pH (3.5) (Ruhanya, 2016) to encourage virus adsorption to filters. Electropositive filters do not need conditioning and operate at near neutral pH (Shi et al., 2017). Large volumes of water are passed through the filter, allowing viruses to adsorb to the filter media. Viruses are then eluted from the filters by an eluting solution. The most widely used eluting solution for viruses is beef extract. Other solutions include glycine and skim milk (Cantalupo et al., 2011; Shi et al., 2017). Viruses from sludge and biosolid samples are eluted according to ASTM D4994-89 (ASTM, 2014). Common secondary concentration methods for environmental water samples include organic flocculation, membrane filtration, tangential flow filtration (TFF), ultracentrifugation, density-dependent centrifugation, polyethylene glycol (PEG) precipitation, and nuclease treatment (Table 1.1). Several other concentration methods for virus detection have also been explored (Hjelmsø et al., 2017). Additionally, concentration techniques may vary depending on the targeted virus and water characteristics (Ahmed et al., 2015; Deboosere et al., 2011; Katayama et al., 2002). In addition to laboratory concentration methods, researchers may choose to implement in silico enrichment techniques post metagenomic sequencing. Removing nontargeted sequences from a metagenome can improve the accuracy of downstream virus classification processes. Decontamination of nontargeted sequences increases the likelihood of detecting less-dominant viruses in metagenomes and reduces computational resources and ambiguity during metagenomic assembly. The use of genetic markers such as the 16S rRNA gene in bacteria, and mapping sequences to nontargeted organism databases (i.e., human or prokaryotic, among others) have been employed in recent virome studies (Cantalupo et al., 2011; Djikeng et al., 2009; Li et al., 2015; Motlagh et al., 2017). Despite the benefits of in silico enrichment, metagenomic data sets can be reduced up to 90% (Rose et al., 2016), resulting in a loss of crucial biological information and a reduction in the comprehensiveness of the study. Generally, annotated metagenomes contain a significant portion of prokaryotic genomes due to the increased likelihood of their larger genomes being sequenced over smaller viral genomes, the presence of prophages annotated as their host genomes, and the limited number of viral reference sequences available in public databases (Fancello et al., 2012; Tamaki et al., 2012). Therefore, laboratory concentration techniques for viral diversity studies are favored over in silico enrichment. A recent study on the diversity of viruses in polluted waters in Uganda found virus-related sequences to comprise as much as 99.79% of affiliated sequences when using the EPA-approved VIRADEL method (O'Brien et al., 2017b). This further suggests that laboratory concentration techniques dictate the quality and richness of postsequenced viromes.

Table 1. 1. Summary of methods for taxonomic classification along with significant findings of viral metagenomic studies in water environments. Phage specific methods are excluded from summary. Various BLAST searches (i.e. BLASTx, BLASTn, tBLASTx) are denoted as BLAST. Percent of human viruses and phages are relative to viral affiliated sequences.

Environmental Media	Virus Concentration Method	Sequencing Platform	Assembler	Aligner/ Annotator	Database	Human Viruses (%)	Phages (%)	Significant Human Virus ^a	Reference
Wastewater	Flocculation with skim milk solution; Ultracentrifugation; Filtration (0.45 µm); Nuclease treatment	Illumina	CLC Genomics	BLAST	NCBI RefSeq (Viral); NCBI Genbank (Viral); UniProt (Viral)	NR	NR	Mamastrovirus	Fernandez-Cassi et al., 2018
Wastewater	Electropositive filtration (NanoCeram); Flocculation with beef extract; Filtration (0.22 µm)	Illumina	IDBA-UD	MetaVir2 (BLAST); Bowtie2	NCBI RefSeq (Viral)	0.53 - 1.21	64.55 - 86.15	Herpesvirus	O'Brien et al., 2017a
Surface water; Wastewater	Electropositive filtration (NanoCeram); Flocculation with beef extract; Filtration (0.22 um)	Illumina	IDBA-UD	BLAST	NCBI RefSeq (Viral)	1.18 - 5.40 ^b	19 - 78	Rotavirus	O'Brien et al., 2017b
Seawater	Polyethersulfone membrane pre-filtration; TTF ^c ; Centrifugal ultrafiltration; CsCl density gradient ultracentrifugation	454/Roche	CLC Genomics	BLAST	NCBI non- redundant protein (Non-viral specific)	NR	NR	NR	Miranda et al., 2016
Ballast water; Harbor water	TFF; PEG ^d ; Chloroform treatment; Filtration (0.45 and 0.22 μm); Nuclease treatment	Illumina	IDBA-UD	BLAST	NCBI RefSeq (Viral)	> 12.6	62.1	Parvoviridae	Kim et al., 2015
Wastewater	PEG; Chloroform treatment; Filtration (0.45 and 0.2 μm); Nuclease treatment	Illumina	Velvet	BLAST	NCBI RefSeq (Viral)	3	67.2	Adenoviridae	Aw et al., 2014
Dairy wastewater	Filtration (0.22 µm)	454/Roche	NR	BLAST	NCBI Genbank (Viral)	NR	13.2 - 23.1	Circoviruses	Alhamlan et al., 2013
Sewage sludge	Flocculation with glycine; Filtration (0.45 μm); PEG	Illumina	Velvet	MG-RAST (BLAST); BLAST	NCBI RefSeq (Viral) – Amended	0.58	NR	Herpesvirus	Bibby and Peccia, 2013
Surface water	Filtration (0.45 µm); PEG; CsCl density gradient ultracentrifugation; Nuclease treatment	454/Roche	Genome Sequence De Novo	MG-RAST (BLAST)	SEED (Non-viral specific)	NR	> 92%	NR	Fancello et al., 2013

Table 1.1 (cont'd)

Wastewater; Activated sludge	Glassfibre pre-filtration; Nitrocellulose filtration (1.2 µm); TFF; CsCl density gradient ultracentrifugation; Nuclease treatment	454/Roche	Newbler	MG-RAST (BLAST)	SEED (Non-viral specific)	NR	16.2 - 45.2	NR	Tamaki et al., 2012
Wastewater	TFF; Sucrose density gradient centrifugation; Nuclease treatment	454/Roche	NR	BLAST	Genbank non- redundant protein and nucleotide (Non-viral specific)	0.66	13.5	Aichi virus	Ng et al., 2012
Class B biosolids	ASTM Method D4994- 89; Filtration (0.45 mm); Nuclease treatment	454/Roche	Newbler	BLAST	NCBI RefSeq (Viral)	<0.1	66.2	Parechovirus	Bibby et al., 2011
Wastewater	Flocculation with skim milk solution; Ultracentrifugation; Nuclease treatment	454/Roche	Phrap	BLAST	NCBI RefSeq (Viral)	5.8	80	Klassevirus	Cantalupo et al., 2011
Surface water	TFF; Ultracentrifugation; Nuclease treatment	454/Roche	Newbler (Hybrid)	BLAST	CAMERA (Non- viral specific)	NR	NR	Banna virus	Djikeng et al., 2009
Potable water; Reclaimed water	TFF; Filtration (0.22 μm); PEG; Chloroform treatment; CsCl density gradient ultracentrifugation; Nuclease treatment	454/Roche	SeqMan	BLAST	Genbank non- redundant protein (Non-viral specific)	4 ^e	26 ^e	Rhinovirus	Rosario et al., 2009

Note: NR = not reported. ^aStudies may have detected more human viruses than what is listed in this table. Viruses listed signify significant finding. ^bPercentage represents vertebrate host (may not all be of human origin). ^cTangential flow filtration (TFF). ^dPolyethylene glycol (PEG). ^ePercentages reflect results in reclaimed water

5. Quality Analysis of Viral-Related Sequence Reads

Sequencing platforms produce massive amounts of short DNA fragments (reads). The first goal for many viral metagenomic studies is to accurately assemble and combine DNA fragments, thus reconstructing the whole genome. The quality of raw reads directly affects the assembly of these fragments into longer sequences or the direct mapping of sequence reads to reference databases. Therefore, prior to assembling, it is common practice to trim low-quality bases and sequencing artifacts as well as remove duplicate and short reads (Kunin et al., 2008).

The quality of a sequenced nucleotide base is generally evaluated by its Phred quality score, which is determined by the sequencing platform used to generate reads. A Phred quality score is a quality parameter adopted by most quality control and analysis tools for read trimming and filtering (Ruffalo et al., 2011). Quality trimming may be applied over the entire read or within a specified window of bases. Phred score thresholds ranging from 15 to 30 are commonly set in viral metagenomic studies (i.e., only bases or reads \geq Q are used in the analysis) (Aw et al., 2014; Bibby and Peccia, 2013; Fernandez-Cassi et al., 2018; Kim et al., 2015; Miranda et al., 2016), where a Phred score of 20 is considered good quality (Lapidus, 2009).

In addition to quality trimming, removal of sequencing artifacts is another critical step to ensure good-quality reads for assembly and alignment. Adaptors, synthetic oligonucleotides, are added to a sample's DNA to condition the DNA for sequencing. Removal of adaptors postsequencing is necessary to reduce contamination from these synthetic sequences. Adaptor trimming is typically carried out at the 3' end of a sequence (Aw et al., 2014; Bibby and Peccia, 2013; O'Brien et al., 2017b). Because sequencing is carried out from the 5' to 3' ends of the DNA fragment, fragments
shorter than the sequencing length can cause the 3' adaptor to lie within the sequencing region (Bolger et al., 2014). Adaptor contamination may be low depending on the type of data sequenced, sequencing platform, and sequencing length. However, the presence of adaptors in viromes can significantly increase the number of unassigned reads during read alignment and annotation, and therefore, it is best practice to remove adaptors prior to assembly. Another quality control practice includes dereplication, the removal of redundant reads (Bibby et al., 2011; Bzhalava and Dillner, 2013; Cantalupo et al., 2011; Fernandez-Cassi et al., 2018; Miranda et al., 2016; Tamaki et al., 2012). Replicate filtering depends on the sequencing platform used to generate reads. Illumina and 454 (Roche, Basel, Switzerland) pyrosequencing platforms are widely used in viral metagenomic studies (Table 1.1) with the latter being recently discontinued. The generation of artificial replicates in 454 sequencing technologies facilitates the need for replicate filtering, as opposed to sequences generated from Illumina platforms (Gomez-Alvarez et al., 2009). In-depth comparisons between 454 and Illumina sequencing technologies have been previously reviewed (Liu et al., 2012; Luo et al., 2012). Additionally, removal of reads significantly shorter than the specified read length is a typical quality filtering approach in viral meteganomic studies. These reads are often less informative, increase computing time, and can results in assembly errors (White et al., 2017). Read length thresholds depend on read length and application of the study.

6. Assembly of Viromes

Assembly is the process of assembling sequence reads into longer contiguous sequences (contigs). Although not required for taxonomic classification, it results in a decrease in data volume, which reduces time and computational resources. Assembling reads also improves the accuracy of alignment processes because longer sequences contain more genetic information. For these reasons, assembly is a standard practice in viral metagenomic studies. The process of assembling reads is highly dependent on the type of sequencing platform used and reads generated. There are two main approaches to assembly, reference-based and de novo assembly. Reference-based assembly involves the use of a backbone genome in which unknown sequences can be referenced to and assembled. De novo assembly uses similarities between reads to obtain a consensus sequence or suggested whole genome. De novo assembly is most commonly used in metagenomic analyses because assigning backbone genomes to an unknown complex community is nearly impossible. Optimized de novo assembly practices for viral genome reconstruction include combined and iterative exhaustive assembly approaches. These approaches involve combining contigs produced from multiple assembly platforms or repeated assembly of the same set of contigs and unassembled reads until no other contigs can be formed. These methods have been known to improve accuracy and recover a large portion of the metagenome (Djikeng et al., 2009; Smits et al., 2014; White et al., 2017). The accuracy of assembly in viral metagenomic studies can be distorted by sequencing artifacts, volume of data, presence of nontargeted organisms, and length and quality of read (Lapidus, 2009; Rose et al., 2016). Short contigs may misrepresent virus abundance in metagenomes because they can create bias when assessing population estimates (Roux et al., 2017; Vázquez-Castellanos et al., 2014). Contig length thresholds can be influenced by the application of the study, assembler properties, and length of reads. Discarding contigs less than 100 bp (Fernandez-Cassi et al., 2018) and 200 bp (Aw et al., 2014) have been observed in viral metegenomic studies using Illumina data sets. Removing contigs with lengths less than 200 bp (Rosario et al., 2009) up to 300 bp (Fancello et al., 2013) have been observed in studies using 454 data sets.

7. Alignment and Taxonomic Classification of Viromes

There are two main approaches to taxonomic classification in metagenomic studies: (1) composition-based approaches, and (2) similarity-based approaches. The latter is most widely used in viral metagenomic studies because few tools exist that are trained on viral genomes. Similaritybased searches involve aligning assembled contigs or raw reads against reference genomes or genes. In viral studies, aligning contigs is considered best practice considering the presence of highly divergent sequences and shorter genomes. However, O'Brien et al. 2017a aligned short reads, also known as fragment alignment (Sharpton, 2014), against the RefSeq viral database as a corroboratory detection method for assessing virus diversity in three sewage-treatment plants. Fragment alignment detected herpesvirus and adenovirus in sewage samples; this was similar to results obtained from aligning contigs using Metavir annotation platform. Various short-read aligners such as Burrows-Wheeler alignment tool (BWA), and Bowtie2 are designed for aligning a large number of short reads to large reference data sets. Generally, in viral metagenomic studies, this approach is most commonly used for evaluating contig abundance and in silico decontamination of nontargeted organisms (Bibby and Peccia, 2013; Cantalupo et al., 2011; Li et al., 2015).

Despite the significant number of similarity-based aligners available, the basic local alignment search tool (BLAST) is the most widely used alignment tool in viral metagenomic studies. The translated nucleotide query against translated nucleotide database BLAST (tBLASTx) algorithm with an E-value of 10^{-3} is said to be more accurate when identifying viral pathogens compared to the translated nucleotide query against protein database BLAST (BLASTx) and nucleotide BLAST (BLASTn) searches (Bibby et al., 2011). Hence, a number of studies have adopted this approach

for detecting less-abundant human pathogens in viromes (Alhamlan et al., 2013; Aw et al., 2014; Bibby et al., 2011; Bibby and Peccia, 2013; Cantalupo et al., 2011). However, a number of viral diversity studies have performed BLASTx alignment with an E-value of 10^{-5} (Djikeng et al., 2009; Fernandez-Cassi et al., 2018; Kim et al., 2015; Miranda et al., 2016). Regardless, Kunin et al. 2008 suggested that protein searches increase the sensitivity of alignment, therefore increasing the accuracy of annotated sequences. This may be especially true for viruses because protein searches are more likely to identify divergent sequences based on conserved regions (Fancello et al., 2012).

The type of reference database also plays a role in the accuracy of alignment. Bibby et al. 2011 found that viral specific databases can improve detection compared with reference databases containing organisms across multiple groups. Aligning genomes against target-specific databases are less computationally demanding due to the reduction in data and decreases ambiguity during alignment. The National Center for Biotechnology Information (NCBI) Viral RefSeq database contains validated nonredundant sequences of viral origin (O'Leary et al., 2016) and has become a standard database in viral studies (Table 1.1). NCBI RefSeq has made strides to standardize reference sequences among public databases to allow for more conclusive comparative studies, less confusion between various annotation platforms, and increased accuracy when annotating viral sequences (O'Leary et al., 2016). Other databases include the NCBI nonredundant database, the Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA) database (Sun et al., 2011), and SEED (Overbeek et al., 2014). These databases consist of publicly available sequences belonging to organisms from multiple taxonomic groups.

Several factors influence the alignment of sequences against reference databases. In particular, contamination can distort viral diversity when assessing whole communities. The presence of nontargeted organisms and dominant populations within viromes create difficulties when detecting viral pathogens in metagenomes. Bacteriophages are the most abundant entities on Earth and usually constitute a large portion of affiliated viruses in metagenomic studies (Table 1.1). This imposes further complications when aiming to detect human pathogens in viromes and even more so when viruses are contaminated with bacterial genomes. Tamaki et al. 2012 studied the diversity and functions of DNA viruses in major stages of a domestic sewage-treatment plant in Singapore using metagenomic analysis. Despite performing virus enrichment prior to sequencing, affiliated sequences contained 79%–84% of bacteria related sequences, and 13%–18.5% of viral sequences of which none were affiliated with human viruses. High numbers of bacteria affiliated sequences in viromes may be related to inadequate virus enrichment, horizontal gene transfer by bacteriophages, and prophages annotated as their host genomes (Cantalupo et al., 2011; Tamaki et al., 2012). The low abundance and relatively short genomes of human viruses are of particular concern in metagenomics studies because they may fail to assemble or go undiscovered during taxonomic classification.

8. Limitations of Next-Generation Sequencing for Viral Diversity and Pathogen Detection

Despite its achievements, NGS is not without limitations. The chance of a virus' genome being sequenced in a metagenome is a result of its size (Bibby et al., 2011; Fancello et al., 2012). Viral pathogen genomes are generally small in nature and less abundant, especially RNA viruses, which creates a universal challenge when detecting these organisms and determining their abundance in

metagenomes (Marz et al., 2014). Sequencing platforms such as Illumina try to overcome this limitation by producing short reads at greater sequencing depths to increase the chance of detecting less-abundant genomes (Nieuwenhuijse and Koopmans, 2017). Consequently, the application of the study should be taken into consideration when selecting the appropriate sequencing platform. Similarly, web-based annotation platforms like Viral Informatics Resource for Metagenome Exploration (VIROME) (Wommack et al., 2012) and Metavir (Roux et al., 2014) are designed to increase virus annotation and improve comparison of viral communities across different environments. A comprehensive review of bioinformatics tools used for analyzing viral sequences has been discussed elsewhere (Nooij et al., 2018; Sharma et al., 2015).

Despite methods to overcome sequencing bias, the limited number of available reference sequences for viral genomes remains a challenge across all annotation platforms. Because viruses lack a universal marker, the detection of viruses in metagenomes is typically done through assembling of contigs and comparisons against previously identified sequences. The highly divergent nature of viruses makes it difficult to maintain well-updated reference databases, resulting in a large number of unaffiliated sequences in viral studies. Moreover, NGS sequencing and bioinformatics tools are prone to a number of errors, and therefore molecular methods such as qPCR are needed to validate findings from NGS analyses. The poor quantitative nature of NGS and difficulty in determining pathogen infectivity further promotes the need for confirmatory methods tailored to the application of the study (Goldberg et al., 2015). Regardless of its limitations, NGS has greatly advanced the current state of knowledge in viral diversity studies with powerful bioinformatics tools to combat sequencing limitations for improved accuracy.

9. Summary and Conclusions

Traditional molecular techniques are limited to identifying well known viruses and assessing microbial environments only through the lens of targeted organisms. Next-generation sequencing enables researchers to obtain a rapid, unbiased view of viral communities with the possibility of discovering novel pathogens, thus contributing a great deal of knowledge in the field of environmental virology. Sequencing and computational errors in viral metagenomics, as well as the poor quantitative nature of NGS, calls for traditional confirmatory methods such as cell cultures and qPCR. Despite the limitations of viral sequencing, bioinformatics tools are continuously developing to provide a deeper understanding of the fate and dynamics of viruses in water environments. With that, standard approaches such as virus enrichment, quality control, assembly, and alignment approaches have been adopted across multiple applications to facilitate a clear understanding of viromes across different ecosystems. Here, the authors have reviewed standard approaches as well as suggested best practices for viral diversity studies in water environments using metagenomics. These techniques, paired with confirmatory methods, make NGS an effective method of detection in environmental studies with the potential to influence environmental infrastructures and policies in the aim of reducing the risk of human exposure to viral pathogens.

REFERENCES

REFERENCES

- Ahmed, W., Harwood, V.J., Gyawali, P., Sidhu, J.P.S., Toze, S., 2015. Comparison of concentration methods for quantitative detection of sewage-associated viral markers in environmental waters. Appl. Environ. Microbiol.
- Alhamlan, F.S., Ederer, M.M., Brown, C.J., Coats, E.R., Crawford, R.L., 2013. Metagenomicsbased analysis of viral communities in dairy lagoon wastewater. J. Microbiol. Methods.
- Anderson, S., 1981. Shotgun DNA sequencing using cloned DNase I-generated fragments. Nucleic Acids Res.
- ASTM, 2014. Standard practice for recovery of viruses from wastewater sludges. In: ASTM D4994-89. West Conshohocken, PA.
- Aw, T.G., Howe, A., Rose, J.B., 2014. Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater. J. Virol. Methods.
- Bibby, K., Peccia, J., 2013. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. Environ. Sci. Technol.
- Bibby, K., Viau, E., Peccia, J., 2011. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. Lett. Appl. Microbiol.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics.
- Bzhalava, D., Dillner, J., 2013. Bioinformatics for Viral Metagenomics. J. Data Mining Genomics Proteomics 4.
- Cantalupo, P.G., Calgua, B., Zhao, G., Hundesa, A., Wier, A.D., Katz, J.P., Grabe, M., Hendrix, R.W., Girones, R., Wang, D., Pipas, J.M., 2011. Raw sewage harbors diverse viral populations. MBio 2, 1–11.
- Deboosere, N., Horm, S.V., Pinon, A., Gachet, J., Coldefy, C., Buchy, P., Vialette, M., 2011. Development and validation of a concentration method for the detection of influenza a viruses from large volumes of surface water. Appl. Environ. Microbiol. 77, 3802–3808.
- Djikeng, A., Kuzmickas, R., Anderson, N.G., Spiro, D.J., 2009. Metagenomic analysis of RNA viruses in a fresh water lake. PLoS One 4.
- Doan, T., Wilson, M.R., Crawford, E.D., Chow, E.D., Khan, L.M., Knopp, K.A., O'Donovan, B.D., Xia, D., Hacker, J.K., Stewart, J.M., Gonzales, J.A., Acharya, N.R., DeRisi, J.L., 2016. Illuminating uveitis: Metagenomic deep sequencing identifies common and rare pathogens. Genome Med. 8.

- Dulbecco, R., Vogt, M., 1954. Plaque formation and isolation of pure lines with poliomyelitis viruses. J. Exp. Med. 99, 167–182.
- Fancello, L., Raoult, D., Desnues, C., 2012. Computational tools for viral metagenomics and their application in clinical research. Virology 434, 162–174.
- Fancello, L., Trape, S., Robert, C., Boyer, M., Popgeorgiev, N., Raoult, D., Desnues, C., 2013. Viruses in the desert: A metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. ISME J. 7, 359–369.
- Fernandez-Cassi, X., Timoneda, N., Martínez-Puchol, S., Rusiñol, M., Rodriguez-Manzano, J., Figuerola, N., Bofill-Mas, S., Abril, J.F., Girones, R., 2018. Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. Sci. Total Environ. 618, 870–880.
- Gardner, R.C., Howarth, A.J., Hahn, P., Brown-Luedi, M., Shepherd, R.J., Messing, J., 1981. The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing. Nucleic Acids Res. 9, 2871–2888.
- Goldberg, B., Sichtig, H., Geyer, C., Ledeboer, N., Weinstock, G.M., 2015. Making the leap from research laboratory to clinic: Challenges and opportunities for next-generation sequencing in infectious disease diagnostics. MBio 6.
- Gomez-Alvarez, V., Teal, T.K., Schmidt, T.M., 2009. Systematic artifacts in metagenomes from complex microbial communities. ISME J. 3, 1314–1317.
- Greninger, A.L., Naccache, S.N., Federman, S., Yu, G., Mbala, P., Bres, V., Stryke, D., Bouquet, J., Somasekar, S., Linnen, J.M., Dodd, R., Mulembakani, P., Schneider, B.S., Muyembe-Tamfum, J.J., Stramer, S.L., Chiu, C.Y., 2015. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 7.
- Hall, R.J., Wang, J., Todd, A.K., Bissielo, A.B., Yen, S., Strydom, H., Moore, N.E., Ren, X., Huang, Q.S., Carter, P.E., Peacey, M., 2014. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. J. Virol. Methods 195, 194–204.
- Heather, J.M., Chain, B., 2016. The sequence of sequencers: The history of sequencing DNA. Genomics 107, 1–8.
- Hjelmsø, M.H., Hellmér, M., Fernandez-Cassi, X., Timoneda, N., Lukjancenko, O., Seidel, M., Elsässer, D., Aarestrup, F.M., Löfström, C., Bofill-Mas, S., Abril, J.F., Girones, R., Schultz, A.C., 2017. Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing. PLoS One 12, 1–17.
- Katayama, H., Shimasaki, A., Ohgaki, S., 2002. Development of a virus concentration method and its application to detection of enterovirus and Norwalk virus from coastal seawater.

Appl. Environ. Microbiol. 68, 1033–1039.

- Kim, Y., Aw, T.G., Teal, T.K., Rose, J.B., 2015. Metagenomic Investigation of Viral Communities in Ballast Water. Environ. Sci. Technol. 49, 8396–8407.
- Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., Hugenholtz, P., 2008. A Bioinformatician's Guide to Metagenomics. Microbiol. Mol. Biol. Rev. 72, 557–578.
- Lapidus, A.L., 2009. Genome sequence databases: Sequencing and assembly. In: Encyclopedia of Microbiology.
- Leland, D.S., Ginocchio, C.C., 2007. Role of cell culture for virus detection in the age of technology. Clin. Microbiol. Rev. 20, 49–78.
- Li, L., Deng, X., Mee, E.T., Collot-Teixeira, S., Anderson, R., Schepelmann, S., Minor, P.D., Delwart, E., 2015. Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. J. Virol. Methods 213, 139–146.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M., 2012. Comparison of next-generation sequencing systems. J. Biomed. Biotechnol. 2012.
- Luo, C., Tsementzi, D., Kyrpides, N., Read, T., Konstantinidis, K.T., 2012. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS One 7.
- Lysholm, F., Wetterbom, A., Lindau, C., Darban, H., Bjerkner, A., Fahlander, K., Lindberg, A.M., Persson, B., Allander, T., Andersson, B., 2012. Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. PLoS One 7.
- Marz, M., Beerenwinkel, N., Drosten, C., Fricke, M., Frishman, D., Hofacker, I.L., Hoffmann, D., Middendorf, M., Rattei, T., Stadler, P.F., Töpfer, A., 2014. Challenges in RNA virus bioinformatics. Bioinformatics 30, 1793–1799.
- Maxam, A.M., Gilbert, W., 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U. S. A. 74, 560–564.
- Miranda, J.A., Culley, A.I., Schvarcz, C.R., Steward, G.F., 2016. RNA viruses as major contributors to Antarctic virioplankton. Environ. Microbiol. 18, 3714–3727.
- Motlagh, A.M., Bhattacharjee, A.S., Coutinho, F.H., Dutilh, B.E., Casjens, S.R., Goel, R.K., 2017. Insights of phage-host interaction in hypersaline ecosystem through metagenomics analyses. Front. Microbiol. 8, 1–15.
- Ng, T.F.F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B.S., Wommack, K.E., Delwart, E., 2012. High variety of known and new RNA and DNA

viruses of diverse origins in untreated sewage. J. Virol. 86, 12161–12175.

- Nieuwenhuijse, D.F., Koopmans, M.P.G., 2017. Metagenomic sequencing for surveillance of food- and waterborne viral diseases. Front. Microbiol. 8.
- Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., Koopmans, M.P.G., 2018. Overview of virus metagenomic classification methods and their biological applications. Front. Microbiol. 9.
- O'Brien, E., Munir, M., Marsh, T., Heran, M., Lesage, G., Tarabara, V. V., Xagoraraki, I., 2017a. Diversity of DNA viruses in effluents of membrane bioreactors in Traverse City, MI (USA) and La Grande Motte (France). Water Res.
- O'Brien, E., Nakyazze, J., Wu, H., Kiwanuka, N., Cunningham, W., Kaneene, J.B., Xagoraraki, I., 2017b. Viral diversity and abundance in polluted waters in Kampala, Uganda. Water Res.
- O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C.M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W., Maglott, D., Masterson, P., McGarvey, K.M., Murphy, M.R., O'Neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A.R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T.D., Pruitt, K.D., 2016. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44, D733-745.
- Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R., 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42, D206–D214.
- Ramírez-Castillo, F.Y., Loera-Muro, A., Jacques, M., Garneau, P., Avelar-González, F.J., Harel, J., Guerrero-Barrera, A.L., 2015. Waterborne pathogens: Detection methods and challenges. Pathogens 4, 307–334.
- Rosario, K., Nilsson, C., Lim, Y.W., Ruan, Y., Breitbart, M., 2009. Metagenomic analysis of viruses in reclaimed water. Environ. Microbiol. 11, 2806–2820.
- Rose, R., Constantinides, B., Tapinos, A., Robertson, D.L., Prosperi, M., 2016. Challenges in the analysis of viral metagenomes. Virus Evol. 2.
- Roux, S., Emerson, J.B., Eloe-Fadrosh, E.A., Sullivan, M.B., 2017. Benchmarking viromics: An in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ.

- 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.
- Ruffalo, M., Laframboise, T., Koyutürk, M., 2011. Comparative analysis of algorithms for nextgeneration sequencing read alignment. Bioinformatics 27, 2790–2796.
- Ruhanya, V., 2016. Adsorption-Elution Techniques and Molecular Detection of Enteric Viruses from Water. J. Hum. Virol. Retrovirology 3.
- Sanger, F., Donelson, J.E., Coulson, A.R., Kössel, H., Fischer, D., 1973. Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage fl DNA. Proc. Natl. Acad. Sci. U. S. A. 70, 1209–1213.
- Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467.
- Sharma, D., Priyadarshini, P., Vrati, S., 2015. Unraveling the Web of Viroinformatics: Computational Tools and Databases in Virus Research. J. Virol. 89, 1489–1501.
- Sharpton, T.J., 2014. An introduction to the analysis of shotgun metagenomic data. Front. Plant Sci. 5.
- Shi, H., Pasco, E. V., Tarabara, V. V., 2017. Membrane-based methods of virus concentration from water: A review of process parameters and their effects on virus recovery. Environ. Sci. Water Res. Technol. 3, 778–792.
- Smits, S.L., Bodewes, R., Ruiz-Gonzalez, A., Baumgärtner, W., Koopmans, M.P., Osterhaus, A.D.M.E., Schürch, A.C., 2014. Assembly of viral genomes from metagenomes. Front. Microbiol. 5, 1–10.
- Sun, S., Chen, J., Li, W., Altintas, I., Lin, A., Peltier, S., Stocks, K., Allen, E.E., Ellisman, M., Grethe, J., Wooley, J., 2011. Community cyberinfrastructure for advanced microbial ecology research and analysis: The CAMERA resource. Nucleic Acids Res. 39, D546– D551.
- Tamaki, H., Zhang, R., Angly, F.E., Nakamura, S., Hong, P.Y., Yasunaga, T., Kamagata, Y., Liu, W.T., 2012. Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. Environ. Microbiol. 14, 441–452.
- Taylor, M.W., 2014. A History of Cell Culture. In: Viruses and Man: A History of Interactions. pp. 41–52.
- Towner, J.S., Sealy, T.K., Khristova, M.L., Albariño, C.G., Conlan, S., Reeder, S.A., Quan, P.L., Lipkin, W.I., Downing, R., Tappero, J.W., Okware, S., Lutwama, J., Bakamutumaho, B., Kayiwa, J., Comer, J.A., Rollin, P.E., Ksiazek, T.G., Nichol, S.T., 2008. Newly discovered Ebola virus associated with hemorrhagic fever outbreak in Uganda. PLoS Pathog. 4.

- U.S. EPA, 2001. Manual of Methods for Virology (Chapter 14).
- Vázquez-Castellanos, J.F., García-López, R., Pérez-Brocal, V., Pignatelli, M., Moya, A., 2014. Comparison of different assembly and annotation tools on analysis of simulated viral metagenomic communities in the gut. BMC Genomics 15.
- White, D.J., Wang, J., Hall, R.J., 2017. Assessing the impact of assemblers on virus detection in a de novo metagenomic analysis pipeline. J. Comput. Biol. 24, 874–881.
- Wigginton, K.R., Ye, Y., Ellenberg, R.M., 2015. Emerging investigators series: The source and fate of pandemic viruses in the urban water cycle. Environ. Sci. Water Res. Technol. 1, 735–746.
- Wommack, E.K., Bhavsar, J., Polson, S.W., Chen, J., Dumas, M., Srinivasiah, S., Furman, M., Jamindar, S., Nasko, D.J., 2012. VIROME: A standard operating procedure for analysis of viral metagenome sequences. Stand. Genomic Sci. 6, 427–439.
- Wu, R., Taylor, E., 1971. Nucleotide sequence analysis of DNA. II. Complete nucleotide sequence of the cohesive ends of bacteriophage λ DNA. J. Mol. Biol. 57, 491–511.
- Xagoraraki, I., Yin, Z., Svambayev, Z., 2014. Fate of viruses in water systems. J. Environ. Eng. (United States) 140.

CHAPTER 2: EARLY DETECTION OF A HEPATITIS OUTBREAK IN AN URBAN COMMUNITY USING WASTEWATER-BASED EPIDEMIOLOGY

Submitted in part for publication: Camille McCall, Huiyun Wu, Brijen Miyani, Evan O'Brien, William Cunningham, Irene Xagoraraki

Abstract

Early detection methods for viral disease outbreaks can advance public health responses to new or existing viral threats and reduce the spread of disease. This study seeks to employ wastewaterbased epidemiology for early detection of hepatitis A outbreaks in urban communities. Quantitative PCR was implemented on 54 untreated wastewater samples collected during the peak of a hepatitis A outbreak and 58 samples collected post outbreak or sporadic case conditions. An early detection window of 7-9 days was established based a mechanistic model constructed given clinical data and virus infection characteristics. Correlation and multiple linear regression analyses were applied to determine the influence of several factors, namely, population, precipitation, sampling conditions, and the number of disease cases on hepatitis A virus (HAV) concentrations in wastewater during both conditions. Furthermore, samples collected during peak outbreak conditions were subjected to next generation sequencing and metagenomic analysis to identify viral hepatitis types circulating in community sewage.

Average concentrations of HAV in wastewater per sampling date range between 1.71×10^6 – 5.19×10^7 copies/L for peak outbreak conditions and 2.57×10^4 – 3.19×10^6 copies/L for sporadic case conditions. HAV loads in wastewater were strongly correlated to the number of disease cases

during peak outbreak conditions. There was a significant drop in HAV concentration in wastewater following a significant decline in cases during sporadic case conditions with no significant temporal correlation between disease cases and concentration. According to the multiple linear regression model, the number of cases had a strong significant affect on HAV concentrations following precipitation and sampling location. Metagenomic analysis identified HAV, hepatitis E virus, and hepatitis C virus in samples collected during outbreak conditions. This study demonstrates the potential use of wastewater-based epidemiology for detection of hepatitis A outbreaks approximately 7 to 9 days before cases are reported to health care facilities and routine environmental monitoring of viral hepatitis in communities.

1. Introduction

Infectious viral outbreaks can cause uncontrollable negative effects especially in densely populated areas. Identification and rapid detection are critical for effective management and prevention of outbreaks. Wastewater-based epidemiology (WBE) is a promising methodology for early detection of viral outbreaks at a population level (O'Brien and Xagoraraki, 2019; Xagoraraki and O'Brien, 2020). Untreated wastewater harbors a wealth of information about the community in the sewage catchment area. Centralized wastewater treatment facilities have the capacity to collect wastewater from thousands or millions of inhabitants per day revealing valuable information about the serviced population and potentially providing early signs of viral outbreaks. Specifically, hepatitis A virus (HAV) has caused significant outbreaks worldwide. HAV is a non-enveloped single-stranded RNA virus belonging to the *Picornaviridae* family. HAV is an enteric virus transmitted through the fecal-to-oral route and spreads via person-to-person or contaminated food and water (Lemon et al., 2018). The burden of hepatitis A outbreaks has had a significant impact

on communities and healthcare infrastructures (Snyder et al., 2019). There are approximately 1.5 million cases of hepatitis A reported annually. According to the World health organization, HAV infections resulted in 11,000 deaths in 2015 (WHO, 2017).

Previous studies have evaluated wastewater surveillance of hepatitis A in communities (Bisseux et al., 2018; Chen et al., 2019; Gharbi-khelifi et al., 2007; La Rosa et al., 2014; Manor et al., 2017; Yanez et al., 2014) using PCR and comparing the detection rate of positive HAV sewage samples to the incidence of clinical cases reported in the catchment area. However, detection rates were, at times, not correlated with clinical records. Several factors influence the detection of viruses in wastewater including the sensitivity of the method, environmental conditions, and the inherently low levels of viral pathogens in water systems. This is especially a challenge with qualitative (presence/absence) tests. Hellmer et al., 2014 used qPCR to detect HAV in sewage samples in Scandinavia. Findings indicate the potential of wastewater surveillance for early detection of HAV infections in communities but highlights the challenge of correlating concentration and clinical data for viruses when not accounting for disease patterns. Hepatitis E virus concentrations have been reported to potentially fall below detectable limits in wastewater samples depending on daily flow rates and extent of viral presence in the community (Miura et al., 2016). These prior studies provide evidence that there are several factors influencing the usefulness of WBE, particularly, virus disease patterns (i.e. incubation period and time to viral shedding) and magnitude of disease cases within the neighboring population.

Moreover, several studies have explored the presence of viral pathogens in wastewater using next generation sequencing (NGS) and metagenomics (Aw et al., 2014; Fernandez-Cassi et al., 2018;

Ng et al., 2012; O'Brien et al., 2017). With decreases in processing time, the cause and spread of disease can be understood and mitigation strategies can be implemented within hours (Casto et al., 2019; Greninger et al., 2015). This makes broad surveillance of viral pathogens, including viral hepatitis, in wastewater a feasible option for identifying potential health threats within a community before such information reaches local health facilities.

This study aims to evaluate WBE for early detection of a hepatitis A in a large metropolitan area in Michigan during outbreak and sporadic case conditions. The authors used qPCR to evaluate HAV concentrations in untreated wastewater from a large urban municipal wastewater treatment facility during the peak of the 2017 multi-state hepatitis A outbreak and a year after (post-peak conditions) where fewer sporadic cases were reported. Correlations between HAV concentrations and the number of hepatitis A cases reported in the service community were investigated through mechanistic modeling and correlation analysis. Lastly, NGS was performed on samples collected during peak epidemic conditions to investigate the potential utility of WBE for monitoring the circulation of other viral hepatitis types. To our knowledge, the use of a simple mechanistic model, considering the virus infection cycle, to compare the efficacy of WBE during outbreak conditions and sporadic case conditions has yet to be explored.

2. Methods

2.1. Study Area and Wastewater Sample Collection

Wastewater samples were collected from the Water Resource Recovery Facility (WRRF) located in Detroit, Michigan. The Detroit WRRF is one of the larger single site wastewater treatment plants in the U.S. and treats wastewater from an estimated 3 million inhabitants (GLWA, 2018). It has a primary and secondary treatment capacity of 1,700 million gallons per day (MGD) and 930 MGD, respectively, with an average daily flow of 650 MGD. Detroit's WRRF has a combined sewer system, which collects and treats stormwater along with residential, industrial, and commercial waste. It services the three largest counties, by population, in Michigan. These are Wayne, Oakland, and Macomb counties with a residential land use of 43, 55, and 49%, respectively (Jones et al. 2015). The percentage of municipalities serviced by the WRRF in each county is 50%, 52%, and 49% for Wayne, Oakland, and Macomb counties, respectively (Figure 2.1). The remaining municipalities are served by local or decentralized treatment facilities. The WRRF receives wastewater from its service municipalities via three main interceptors (sewers): North Interceptor-East Arm (NI-EA), Detroit River Interceptor (DRI), and Oakwood-Northwest-Wayne County Interceptor (O-NWI) (Figure 2.1).



Figure 2. 1. Detroit Water and Sewerage Department (DWSD) interceptor schematic (A). Detroit Water Resource Recovery Facility (WRRF) service municipalities in Wayne, Oakland, and Macomb counties in Michigan (B). Service municipalities are based on the 2018 Great Lakes Water Authority sewer map for the DWSD (GLWA, 2018). County borders and areas are represented by solid black lines, shaded regions represent service areas.

Untreated wastewater samples were collected at the WRRF from sampling points located at each of the three interceptors. To investigate HAV concentrations during peak hepatitis A outbreak conditions, samples were collected approximately bi-weekly between November 2017 and February 2018 (sampling year one (SY1)) resulting in 6 sampling events (n=54). During sampling year two (SY2), sporadic cases conditions, samples were collected approximately bi-weekly from October 2018 through March 2019 resulting in 9 sampling occasions (n=58). Approximately 9 samples were collected per sampling date however due to operational conditions during SY2, DRI and O-NWI sampling sites were not sampled on all 9 occasions (Table B3.1). Both sampling years may exclude weeks of, or leading up to, major U.S holidays. Viruses were isolated from untreated wastewater using electropositive NanoCeram column filters following the EPA's virus adsorptionelution protocol (U.S. EPA, 2001). Sewage samples were collected in triplicates for each interceptor with average filtered volumes ranging between 24-44 L per interceptor. Each interceptor was sampled with its own filter house, tubing, and vacuum pump to minimize cross contamination. Additionally, 1 L grab samples were collected in triplicates from each sampling site to assess wastewater physio-chemical characteristics (pH, temperature, conductivity) and creatinine concentrations for population estimates on each sampling date. Direct measurements of pH, temperature, and conductivity were taken on-site using the YSI Professional Plus handheld device. Average pH, temperature, and conductivity was 7.2, 13.5°C, and 1031 µS/cm, respectively, for SY1. In turn, variances were 0.1, 15.8 (°C)², and 46,432.4 (µS/cm)². Similar measurements were observed during SY2 with an average pH, temperature, and conductivity of 7.2, 14.0°C, and 967 μ S/cm, respectively. Variances for SY2 were 0.1, 7.5 (°C)², and 101,210.6 (μ S/cm)², respectively. Virus filters were immediately stored on ice and transported to the Environmental

Virology Laboratory at Michigan State University (MSU) and stored in -20°C until further processing. Grab samples were stored on ice and immediately transferred to the lab, preserved at pH 2, and stored in -20°C.

2.2. Sample Processing and Virus Isolation

Following wastewater sampling, NanoCerem cartridge filters were eluted within 24 hours with 1.5% w/v beef extract (0.05 M glycine, pH 9.5) according to the EPA's protocol (U.S. EPA, 2001). In short, filters were eluted with 1 L of beef extract for a total of 2 min. The pH of the solution was adjusted to 3.5 ± 0.1 and flocculated for 30 min before centrifugation at 2500 g for 15 min. Supernatant was discarded and pellets were resuspended in 30 mL of 0.15 M sodium phosphate (pH 9.0-9.5) followed by a second round of centrifugation carried out at 7000 g for 10 min. The supernatant was neutralized (pH ~7.25) and subjected to filtration using to 0.45µm and 0.22µm syringe filters to eliminate bacterial contamination. Nucleic acid extraction was performed on 140 µL of purified virus concentrate using the QIAamp Viral RNA Mini Kit (Qiagen) following the manufacturer's protocol and eluted in 80 µL of elution buffer. RNA was stored at -80°C until further processing.

2.3. Preparation of HAV Standards

HAV was obtained from ATCC for preparation of standard controls. Nucleic acid was extracted as detailed in the previous section and transformed into One Shot TOP10 chemically competent *Escherichia coli* cells using the TOPO Cloning kit (Invitrogen) following the manufacturer's protocol. Plasmid DNA containing cloned HAV was extracted and quantified as previously described (Munir et al., 2011). The protocol detailed in step two of the subsequent section was utilized to prepare a standard curve with 10-fold serial dilutions of positive HAV controls ranging from 10^3 to 10^{10} genome copies/reaction. The standard curve used to estimate HAV concentrations in collected samples obtained a slope and R-squared value of -3.6 and 0.994, respectively.

2.4. Quantitative RT-PCR and Limit of Detection

Quantitative reverse transcription polymerase chain reaction (RT-qPCR) was used to determine HAV concentrations in RNA samples. RT-qPCR was performed on a Mastercycler ep realplex2 (Eppendorf) in 96-well optical plates. HAV was quantified using a two-step RT-qPCR with previously described primers and probe (Jothikumar et al., 2005). Briefly, viral RNA was reverse transcribed using iScript RT-qPCR Supermix (Bio-Rad) according to the manufacturer's protocol. Five microliters of cDNA, negative control, or positive control was transferred to a 15 μ L reaction mix containing a final concentration of 250 nM for each primer, 150 nM of probe, 1× Lightcycler 480 probes master and sterile nuclease free water. All reactions were performed in triplicates with the following amplification conditions: denaturation at 95°C for 15 min, followed by 45 cycles of 95°C for 15 s, 55°C for 20 s, and 72°C for 15 s.

In order to establish the method's limit of detection (LOD), purified nuclease-free water was spiked with 2-fold serial dilutions of HAV positive control ranging between 12.5 copies/reaction and 100 copies/reaction. Ten replicates of each dilution and negative control were analyzed with identical RT-qPCR conditions as described above. The LOD was defined as the lowest copy number belonging to the serial dilution that yielded a positive PCR response in 95% of occurrences

(Burns and Valdivia, 2008). A PCR response was considered positive if it obtained a quantification cycle (Cq) value paired with a sigmoidal amplification curve. A LOD of 100 viral copies/reaction was obtained as observed in an earlier study (Simmons and Xagoraraki, 2011). Non-detectable HAV concentrations in wastewater sample replicates were reported as one-half of the LOD. All HAV concentrations were normalized according to sampling volumes and reported as copies/l.

2.5. Next Generation Sequencing and Metagenomic Analysis

Sequencing analysis was performed on samples collected during SY1 to investigate viral hepatitis types in wastewater during peak hepatitis A outbreak conditions. Purified nucleic acid from each biological replicate was pooled together for a total of 18 samples, which consist of one sample per interceptor for each of the 6 sampling dates. Nucleic acid from each sample was reverse transcribed and subjected to random amplification as previously described (Wang et al., 2003). Eighteen samples of viral cDNA were sent to the Research Technology Support Facility Genomics Core at Michigan State University for whole-genome shotgun sequencing (WGS). The Illumina TruSeq Nano DNA Library Preparation Kit was used for all cDNA samples. Library preparation was performed on a Perkin Elmer Sciclone G3 robot according to the manufacturer's recommendations. This was followed by sequencing on an Illumina HiSeq4000 platform generating 150 bp paired-end reads.

Sequencing reads generated from WGS were processed on a Unix system through the MSU High Performance Computing Center (HPCC). Raw sequences were analyzed for quality using FastQC, a quality control tool for sequencing data (Andrews, 2010). Sequencing adapters and reads with an average quality score below 20 were removed using Trimmomatic (Bolger et al., 2014). Trimmed reads were assembled with IDBA-UD, a short-read de novo sequence aligner for metagenomic data (Peng et al., 2012). Reads were assembled into contigs using an iterative k-mer approach with k-mer sizes ranging between 40 and 120 in increments of 10. Default values were used for the remaining parameters.

Since viral genomes can be difficult to detect in metagenomic datasets, an optimized multialignment approach was used in order to improve alignment and annotation of viral reads. First, contigs were aligned against the Viral RefSeq database using tBLASTx with an E-value of 10^{-3} . This approach has been known to increase human viral discovery in metagenomic datasets (Bibby et al., 2011). Aligned contigs were assigned to the Lowest Common Ancestor (LCA) according to the NCBI's taxonomy with MEGAN's long read algorithm (v. 6.15.0, Huson et al., 2018). The top 10 percent of BLAST alignments with a minimum bit score of 50 and contig coverage of at least 80% were considered in taxonomic analysis. Default values were used for the remaining parameters. Reads assigned to the Riboviria realm were extracted for further analysis of viral hepatitis types. Riboviria is a realm of viruses containing all RNA viruses and viroids that replicate via a RNA template and includes all known human hepatitis types (A, C, D, E, and G (Pegivirus)) except for hepatitis B virus (HBV). HVB is a DNA virus belonging to the Hepadnaviridae family. This family was not detected in sequenced samples. Contigs annotated as Riboviria were aligned with BLASTx with an E-value of 10⁻⁵ against a custom human virus database containing 5,979 human viral proteins in Swiss-Prot database (Boeckmann et al., 2003) including human hepatitis protein sequences. These sequences represented all human viruses in the Swiss-Prot database at the time of retrieval (September 2019).

2.6. Creatinine Analysis

To determine the impact of potential population fluctuations on HAV concentration, creatinine was measured in grab samples collected from each interceptor per sampling date. Creatinine is excreted at a relatively constant rate in the urine of individuals and is therefore considered a potential biomarker for estimating population (Spierto et al., 1997). Liquid chromatography-mass spectrometry was used to assess creatinine concentrations in collected samples as previously noted (Chen et al., 2014). Briefly, creatinine and creatinine (methyl-13C,d3) isotope were purchased from Sigma-Aldrich Corp. Stock creatinine was diluted in methanol to obtain 1, 10, 100, 250, 500, 1000, and 2500 ng/L standard solutions. Internal creatinine standards of 1, 10, 25, 50, 100, and 250 ng/l were prepared by mixing 0.05 mL of creatinine isotope, 0.85 mL of methanol and 0.1 mL of previously prepared standard solution. Solid-phase extraction was carried out on wastewater samples by passing 19.5 mL of sample, 0.05 mL of standard solution, and 1 ml of EDTA through HLB 6 cc vacuum cartridge filters (Oasis) at a rate of 1-3 mL/min. Each filter was eluted with 5 mL of methanol at the same rate. One mL of eluate, methanol, and each chemical standard were loaded into the mass spectrometer. Mobile phase A and B consisted of distilled water with 0.3% folic acid and methanol, respectively.

2.7. Precipitation Data Collection

The Detroit WRRF is a combined sewer system in which stormwater is captured and treated along with wastewater during rainfall events. Climatic changes of this nature are expected to alter daily flows and wastewater compositions. Local precipitation data was utilized to measure the significance of wet weather events on HAV concentrations. Daily precipitation (rainfall and snow

melt) during the study period for all meteorological stations in Wayne, Oakland, and Macomb counties were obtained from the National Oceanic and Atmospheric Administration's (NOAA's) Global Historical Climatology Network (GHCN) database (Menne et al., 2012). Average daily precipitation for each county was determined by averaging the precipitation measurements reported across all stations per day.

2.8. Clinical Data Collection

Disease data for hepatitis A for each service county was obtained from the Michigan Department of Health and Human Services. Weekly counts of confirmed hepatitis A cases were extracted from the Michigan Disease and Surveillance System (MDSS) from 1 January 2017 until 1 June 2019. The MDSS is a communicable disease reporting system used to facilitate coordination and sharing of disease surveillance data among multiple shareholders including healthcare providers and medical laboratories (MDHHS, 2020a). Michigan requires all physicians, healthcare providers, and laboratories to report hepatitis A cases within 24 hours directly to the MDSS or local health department (MDHHS, 2017). Reported cases per county were obtained as weekly aggregate counts and classified as de-identified health information according to the Health Insurance Portability and Accountability Act (HIPPA) Privacy Rule. The number of weekly disease cases are structured according to the CDC Morbidity and Mortality Weekly Report (MMWR), which aggregates the number of cases reported from Sunday to Saturday of each week.

In late 2016 the CDC declared a multi-state hepatitis A outbreak with the primary mode of transmission being person to person (CDC, 2020; Hofmeister et al., 2020). Among the states affected are Michigan resulting in 920 cases, 80% hospitalizations and 30 deaths (CDC, 2020;

MDHHS, 2020b) as of December 2019. The peak of the hepatitis A outbreak in Michigan occurred between August 2017 and December 2017 with a steady decline thereafter (Figure 2.2). Wastewater collection during part of SY1 was considered peak epidemic conditions while collection during SY2 was considered post-peak (sporadic cases) conditions.



Figure 2. 2. Epidemic curve for confirmed hepatitis A cases in Macomb, Oakland, and Wayne counties from January 2017 through April 2019. Number of cases were provided by the MDHHS. Sampling year one (SY1) and sampling year two (SY2) are denoted by dashed lines.

2.9. Mechanistic Modeling for Selection of Hepatitis A Cases

A mechanistic model based on virus incubation period, time of peak viral shedding in feces, wastewater detention time, and sampling frequency was used to determine which hepatitis A cases to associate with each sampling date. A median incubation period of 28 days (Fiore, 2004), and peak viral shedding range of 10-12 days from exposure (CDC, 2015) were considered in the model. Average detention times were determined by estimating the time it takes for wastewater to get from the furthest point in each interceptor to the wastewater treatment plant. Detention times were estimated under normal dry weather conditions using Manning's equations to identify significant lags in wastewater transport. Equations 1 and 2 were used to calculate the slope of energy grade line S (m/m) and the cross-sectional average velocity v (m/s) where n (adimensional) is the coefficient of roughness, Q (m³/s) is the flowrate, and D (m) is the average interceptor diameter (Davis, 2010).

$$S = \frac{10.3n^2Q^2}{D^{16/3}}$$
(1)

$$\mathbf{v} = \left(\frac{0.397}{n}\right) \mathbf{D}^{2/3} \mathbf{S}^{1/2} \tag{2}$$

Majority of the Regional Wastewater Collection System (RWCS) transports wastewater to the WRRF by gravity. A coefficient of roughness of 0.013 was estimated based on pipe material, centrifugally spun concrete (Davis, 2010). Average flowrate and interceptor dimensions are taken from the GLWA 2019-2023 Capital Improvement Plan (CIP) (GLWA, 2018). Flowrates per interceptor are determined based on the percentage of flow coming from each interceptor as specified in the CIP. The detention time in each interceptor was determined by dividing the length of each interceptor by the estimated average flow velocity. Assuming uniform partial flow conditions, average detention times under dry weather conditions were less than one day (4.3-13.7 hours) for each interceptor. Given the long incubation period of HAV and aggregated health data, detention time was considered negligible and excluded from the model.

Figure 2.3 displays the selection window for hepatitis A cases for each sampling date. Wastewater collection was performed on Fridays during SY1 and on Wednesday or Thursday during SY2. Sampling days were used as a reference point for creating the time scale. The authors back-calculated from the reference point to infer the day of exposure for infected persons excreting HAV in feces during the day of sampling. The median incubation period was positioned based on the day of exposure with a range of 15-50 days (Fiore, 2004). A two-week selection window spanning one week before and one week after the median incubation period was selected based on weekly aggregated health data and to account for the wide variation in incubation times. Based on the mechanistic model, a possible early detection period of seven to nine days can be established for hepatitis A outbreaks using WBE. Weeks selected for clinical data are specified in Table 2.1.



Figure 2. 3. Mechanistic model for correlating HAV concentrations in wastewater and hepatitis A cases in the service community. Time scale is in days with one-week increments. EDW = Early Detection Window.

2.10. Statistical Analyses

All statistical analyses were performed in R (R Core Team, 2019). The Performance package (v. 0.4.5) was used to evaluate the distribution of HAV concentrations and the performance of

multiple linear regression analysis unless otherwise stated. The Wilcoxon signed-rank test was used to investigate significant differences in counted or measured variables between sampling periods. Bonferroni's corrected Dunn's nonparametric pairwise test was used to assess significance between sampling dates. Furthermore, the nonparametric Spearman's rank correlation analysis was performed to evaluate the agreement between average HAV concentrations in collected samples for each sampling date and clinical cases selected according to Figure 2.3.

Stepwise linear regression via the MASS package (Venables and Ripley, 2002) was performed to identify significant explanatory factors of HAV concentrations in wastewater samples. Each potential model structure was evaluated using the Akaike information criterion (AIC). Biological replicates were treated as independent measurements to avoid loss of information or obtaining false inferences about the relationship between HAV and the explanatory variables. HAV concentrations were log-transformed to satisfy the assumptions of normality according to the Shapiro-Wilks test and visual inspection of the quantile-quantile (Q-Q) plot. Cook's distance, which is used to measure the influence of data observations, was scaled to 4/n (n = number of observations) and used as a threshold to identify influential HAV concentrations over time (Zuur et al., 2010). Influential observations due to non-detects were removed and concentrations that were quantified during RT-qPCR were inspected for measurement errors.

The preliminary model contained the following explanatory variables: interceptor (INTCEP), creatinine concentration (CRE), precipitation (PRECP), and number of cases (CASES) (Equation 3).

$$\log_{10} \text{HAV} \sim \text{CASES} + \text{CRE} + \text{PRECP} + \text{INTCEP}$$
(3)

Creatinine concentrations were reported as the average for each interceptor per sampling date. Precipitation was reported as the total average precipitation in inches over Macomb, Oakland, and Wayne counties per sampling date. Additionally, sampling year (SY), sampling week (SW), and interactions with SY (CASES and SY, CRE and SY, PRECP and SY) were considered. These main and interaction terms obtained high variance inflation factors signifying multicollinearity and were therefore removed from the model to avoid the risk of type II errors (Zuur et al., 2010). The final model, which was obtained from stepwise regression, was checked for potential influential observations and assumptions of linearity using R's regression diagnostics plots. The significance of heteroscedasticity in residuals was evaluated using the Performance package. P-values < 0.05 were consider statistically significant.

A linear-mixed effects model was also explored using the lme4 package (Bates et al., 2015) to consider variations in HAV due to the nested structure of biological replicates. Several structures of random intercepts were explored including an intercept only model, replicates nested within interceptor, and replicates nested withing sampling year. In all three situations, biological replicates accounted for less than 3% of the variance in HAV concentrations (Intra-class correlation coefficient (ICC) = 0.004-0.027). Therefore, the use of a mixed model would have no significant improvement on model fit (Vajargah and Nikbakht, 2015).

3. Results

3.1. Environmental Surveillance of HAV and Correlation to Clinical Cases

HAV was detected in all interceptors sampled during each sampling date for SY1 and SY2 (Figure 2.4). Average HAV concentrations per interceptor range between $1.05 \times 10^7 - 1.79 \times 10^7$ copies/L and $5.12 \times 10^5 - 1.48 \times 10^6$ copies/L for year one and two, respectively. Average HAV concentrations per sampling date range from $1.71 \times 10^6 - 5.19 \times 10^7$ (median 3.89×10^6) copies/L and $2.57 \times 10^4 - 3.19 \times 10^6$ (median 3.96×10^5) copies/L for SYs 1 and 2, respectively (Figure 2.5).



Figure 2. 4. Boxplots for concentrations of HAV in wastewater samples per interceptor during sampling years one (A) and two (C) along with average concentrations per sampling date during years one (B) and two (D). Median concentrations are denoted with a horizontal line. Due to operational conditions during sampling year two (SY2), the Detroit River Interceptor (DRI) and the Oakwood-Northwest-Wayne County Interceptor (O-NWI) sampling sites were not sampled during weeks where no data is reported.

There was a significant decrease in the number of cases (p < 0.01) and HAV concentrations (p < 0.001) from SY1 to SY2 (Figure 2.5) with the greatest cluster of cases associated with the highest average HAV concentration during SY1 (Table 2.1). Spearman's correlation coefficient showed a strong positive correlation between the number of cases reported and HAV concentrations collected approximately one week prior for SY1 ($\rho = 0.943$, p < 0.05). No significant correlation was observed between cases and concentration for SY2 ($\rho = 0.237$, p > 0.05). It is worth mentioning that viral concentrations obtained from the 17-Nov-17 sampling date for SY1 have a significant influence on clinical data correlations (Figure 2.5). Exclusion of data associated with this sampling date resulted in non-significant associations between concentration and cases (p > 0.05). Following inspection for accuracy, concentrations related to the 17-Nov-17 sampling date were considered critical information and therefore were not omitted.



Figure 2. 5. Temporal correlation between selected reported hepatitis A cases in service counties and average measured HAV concentrations in wastewater samples collected during sampling years one (A) and two (B). Error bars represent the standard error of measured concentrations for each date.

Sampling Year	Sampling Week (DD-Month-YY)	Average HAV (copies/L)	Average Creatinine (mg/L)	Precipitation (inches)	No. Of Cases	Clinical Weeks Start Date: End Date (DD-Month-YY)	CDC MMWR Weeks
SY1	17-Nov-17	5.19E+07	295	0.02	29	26-Nov-17: 9-Dec-17	48, 49
SY1	1-Dec-17	4.81E+06	34	0.29	21	10-Dec-17: 23-Dec-17	50, 51
SY1	14-Dec-17	1.83E+07	148	1.14	20	24-Dec-17: 6-Jan-18	52, 1
SY1	19-Jan-18	2.96E+06	208	0.00	12	28-Jan-18: 10-Feb-18	5,6
SY1	2-Feb-18	2.40E+06	8	0.00	9	11-Feb-18: 24-Feb-18	7, 8
SY1	16-Feb-18	1.71E+06	2	0.07	7	25-Feb-18: 10-Mar-18	9, 10
SY2	17-Oct-18	1.28E+06	70	0.00	0	28-Oct-18: 10-Nov-18	44, 45
SY2	31-Oct-18	3.19E+06	619	1.46	1	11-Nov-18: 24-Nov-18	46, 47
SY2	28-Nov-18	9.33E+05	278	0.00	4	9-Dec-18: 22-Dec-18	50, 51
SY2	12-Dec-18	1.65E+06	279	0.02	1	23-Dec-18: 5-Jan-19	52, 1
SY2	17-Jan-19	1.70E+05	559	0.00	0	27-Jan-19: 9-Feb-19	5,6
SY2	7-Feb-19	2.57E+04	93	0.23	1	17-Feb-19: 2-Mar-19	8, 9
SY2	14-Feb-19	3.96E+05	224	0.53	1	24-Feb-19: 9-Mar-19	9, 10
SY2	28-Feb-19	1.70E+05	615	0.03	0	10-Mar-19: 23-Mar-19	11, 12
SY2	14-Mar-19	6.15E+04	883	0.24	0	24-Mar-19: 6-Apr-19	13, 14

Table 2. 1. Average measured and collected environmental and disease data during for each sampling date during study period.

3.2. Effects of Population, Precipitation, and Disease on HAV Concentrations in Wastewater Samples

A linear regression analysis was performed to identify environmental factors influencing HAV concentrations in wastewater samples, namely, sampling site, creatinine concentrations, precipitation, and number of reported hepatitis A cases in the service community. Average creatinine concentrations were significantly higher (p < 0.01) during the second sampling year as compared to the first sampling year (Table 2.1). Post hoc pairwise test revealed no significant difference in creatinine concentrations between sampling dates within the corresponding sampling year. Total precipitation for each sampling day ranged between 0-0.38 inches for SY1 and 0-0.49 inches for SY2. There was no significant difference in precipitation between sampling years (p > 0.05) or between sampling dates (p > 0.05).

Stepwise linear regression found significant differences in HAV concentrations between the NI-EA and DRI sampling sites (β = -1.11, p < 0.05). Moreover, precipitation (β = 0.92, p < 0.05) and the number of cases (β = 0.13, p < 0.001) had a significant effect on HAV concentrations in collected samples. According to stepwise regression analysis, creatinine had no significant effect on variations in HAV concentrations. Diagnostics plots and performance statistics revealed no significant deviation from homoscedasticity (p > 0.05) or normality (Figure A.1). The final model was statistically significant (F = 14.71, p < 0.0001) with an R² value of 0.34. Hence, the overall model accounts for approximately 34% of the variation in HAV concentrations in collected samples with an average residual standard error of ± 1.9 log copies/L (Figure A.2).

3.3. Metagenomic Analysis of Viral Hepatitis

A total of 624.4 million reads were obtained from Illumina sequencing and subject to quality trimming resulting in 595.2 million reads. The proportion of contigs assigned to viral taxonomic groups range between 72-83%. Approximately 0.49% of contigs were assigned to the Riboviria group. There were 8 human viral families detected within the Riboviria realm including families containing hepatitis C virus (HCV) and human pegivirus (HPgV) (*Flaviviridae*), hepatitis E virus (HEV) (*Hepeviridae*), and HAV (*Picornaviridae*) (Figure A.3). Identification at the genus level suggested the presence of three viral hepatitis types, namely, HAV, HEV, and HCV detected in 100%, 72%, and 11% of sequenced samples, respectively. HAV virus obtained a greater relative abundance compared to HEV and HCV in all samples (Figure 2.6).



Figure 2. 6. Relative abundance of contigs assigned to human associated viral hepatitis protein sequences from custom Swiss-Prot database.
4. Discussion

4.1. Environmental Surveillance of HAV and Connection to Clinical Data

With the premise that domestic wastewater can serve as an environmental indicator of community health, WBE is becoming a powerful tool for surveillance and early detection of viral disease outbreaks. Here, HAV was monitored in wastewater during peak and post-peak outbreak conditions to explore the usefulness of WBE for early detection of hepatitis A outbreaks. Presence of HAV in collected samples were reported and a mechanistic model along with correlation analysis was utilized to investigate associations between viral concentrations in wastewater and clinical data reported approximately one week after sampling.

HAV was detected in all sampling locations per sampling date during peak outbreak and sporadic case conditions. Previous studies have reported negative or low detection rates in wastewater in low endemic regions such as the U.S. in cases where no reports of hepatitis A were circulating in the community. For example, qPCR was performed to survey enteric viruses in a sewage treatment plant in the United Kingdom. HAV went undetected in all samples and there were no reports of hepatitis A cases within the study area during the time of sampling (Farkas et al., 2018a). Furthermore, low detection rates, <10%, were reported in intermediate endemic regions using RT-PCR and nested PCR techniques (Kokkinos et al., 2011). These findings suggest that high detection rates were reported in this study because there was significant HAV circulation among inhabitants within the service community. Despite the significant decline in reported cases during SY2, detection rates were greater than 89% suggesting continued circulation of the virus in the environment or excretion of HAV by persons with asymptomatic infections. Similar observations

were reported in previous studies (Bisseux et al., 2018; La Rosa et al., 2014). However, HAV concentrations were significantly lower during SY2 signifying the change in disease incidence. Zhang and Iacono, 2018 estimated that only 9% of infected individuals were symptomatic during a hepatitis A outbreak at a local elementary school. Fecal surveillance of childcare facilities during hepatitis A outbreaks may help to discriminate between environmental prevalence of HAV and occurrence of subclinical infections within the surrounding community. Monitoring for human activity in wastewater samples can also be a complementary approach to WBE to distinguish between environmental background and human input (Matus et al., 2019).

Although presence of viruses in wastewater may be a good indicator of circulation within the community, detection in wastewater samples are confounded by sample processing regimes and environmental conditions. Previous studies observed low detection rates of HAV in raw wastewater samples despite high clinical incidence in the service community (Gharbi-khelifi et al., 2007; Kamel et al., 2010). Additionally, qualitative WBE could pose challenges in distinguishing between baseline and outbreak conditions in endemic regions where circulation of the virus in the environment is expected. Therefore, appropriate models, as described in the following sections, should be developed. Methods used in this study were able to capture high expected detection rates of HAV in collected samples suggesting the efficiency of the sampling method and sensitivity of quantitative PCR assays.

Albeit there were strong correlations between concentration and cases during SY1, viral concentrations obtained from the 17-Nov-17 sampling date had a significant influence on outcomes from Spearman's correlation analysis. Such occurrence warrants the need more rigorous

sampling, which includes an increase in the frequency and number of samples collected to establish a robust early detection window between presence in wastewater and expected increases in the number of clinical cases. There was no significant correlation between HAV concentrations and cases post outbreak conditions. Findings suggest reduced sensitivity of WBE during sporadic incidences of hepatitis A cases or in the absence of elevated viral concentrations across sampling time points. Nonetheless, given that concentrations obtained from the first sampling date were appropriate, monitoring concentrations during outbreak conditions could aid public health officials in detecting early signs of a hepatitis A outbreaks and predict the magnitude of cases reported seven to nine days after viral footprint in wastewater.

4.2. Factors Influencing HAV Concentrations in Wastewater

Stepwise linear regression was conducted to evaluate the influence of sampling site, creatinine, precipitation, and number of cases on HAV concentrations in collected samples. There was a significant difference in the variation of HAV concentrations between the NI-EA and DRI interceptors. According to Detroit Water and Sewerage Department (DWSD) personnel, the NI-EA interceptor contains a greater proportion of domestic waste as compared to the DRI, which carries more industrial waste. Additionally, the DRI interceptor transports a greater portion of stormwater during wet weather events as compared to the NI-EA and O-NWI interceptors. Therefore, it holds the potential to have a greater dilution effect. According to linear regression analysis, there is a 1.11 ± 0.42 log copies/L decrease in HAV concentration from NI-EA to the DRI sampling sites. Differences between interceptors are less likely attributed to variations in pH, conductivity, or wastewater temperature since these characteristics were consistent across interceptors (Table B2.2 and Table B2.3). Previous studies have also reported negligible effects of

these physio-chemical parameters on viral concentrations (Farkas et al., 2018b; Sidhu et al., 2018). Variations between the DRI and NI-EA interceptors is likely a result of industrial waste or dilution due to stormwater, which may lower viral concentrations in influent wastewater and effect the potential of capturing parallels between viral presence in the community and sewage treatment utilities.

Creatinine was measured in wastewater samples collected from each interceptor per sampling date to capture potentially large fluctuations in population that could be attributed to variations in HAV concentrations. An increase in the number of people served during a given time period may increase HAV concentrations although the incidence in hepatitis A cases remains relatively stable. Albeit there was a significant difference in creatinine between sampling years, there was no significant effect reported on HAV concentrations. Wastewater sampling was generally conducted in the morning (9:00am – 11:00am) during sampling year one and during the late morning through the afternoon (11:00am – 3:00pm) during the second sampling year. This proposes that the difference in creatinine concentrations between sampling years are attributed to differences in human diurnal cycles and less likely due to a large influx of people within the service community. It is important to note that unaccounted for increases in wastewater detention time and microbial compositions in wastewater can significantly accelerate creatinine degradation making it unrepresentative of the sampled population (Thai et al., 2014).

The Detroit WRRF collects and treats both stormwater and wastewater in a combined sewer system. Rainfall or snow melt events can contribute significantly to microbial concentrations in wastewater due to dilution, run-off, or changes in wastewater characteristics (Tolouei et al., 2019).

The linear regression analysis suggests that HAV concentrations increase with increases in precipitation, which is contrary to what the authors expected. Several possible explanations for this occurrence involve operational conditions for WRRF, seasonal impacts, or statistical limitations. During storm events the Detroit WRRF diverts sewage to combined sewer overflow (CSO) retention treatment basins (RTBs), and overflow screening and disinfection facilities (SDFs). Sewage is screened and disinfected in RTBs or SDFs and stored in RTBs. Flows are then transported to the WRRF when normal conditions are restored ((GLWA, 2018). The mixing of disinfected wastewater with untreated sewage after wet weather events can cause false drops in HAV concentrations. Although plausible, such a theory is difficult to prove with the information provided and is outside the scope of this study. Further research is needed with detailed information of the WRRF sewerage system during wet and dry weather to fully delineate the impact of wet weather events on viral concentrations. Moreover, sampling was conducted during fall and winter months for both sampling years, which suggest that snow melt could have a potential effect on viral concentrations. Despite the abovementioned notions, there were no significant differences in precipitation between sampling weeks or sampling years, which indicates that there were similar weather patterns throughout the study period. Lastly, precipitation was considered per sampling date only, which does not all the linear model to evaluate the impact of precipitation an individual interceptor. This is important because the DRI and O-NWI interceptors collect a greater proportion of wastewater than the NI-EA interceptor. This further demonstrates the need for a further study outlining hydrological and microbial changes in wastewater per interceptor during wet weather. Discrepancies with the effect of precipitation on viral load in untreated wastewater were also observed previously (Farkas et al., 2018b). Operational conditions within sewerage network can impose complications when determining clear patterns between viral loads in wastewater and precipitation.

Despite other influences, the number of clinical cases had a strong positive effect on HAV concentrations suggesting the sensitivity of WBE for early detection during outbreak conditions. Sampling site, creatinine, precipitation, and number of cases explained less than half of the variation in HAV concentrations in collected samples. This indicates that other factors influencing HAV in wastewater samples were largely unaccounted for in the model. Sampling and concentration techniques, virus degradation, presence of inhibitors, and predation play an important role in viral concentration patterns in water reservoirs (Kim and Unno, 1996; Pinon and Vialette, 2019; Varughese et al., 2018). Including such factors could increase the explanatory power of linear regression models for viral concentration predictions.

4.3. Metagenomic Analysis of Viral Hepatitis in Wastewater

In order to investigate NGS for monitoring viral hepatitis in wastewater, Illumina sequencing was performed on 18 samples collected during SY1 to assess viral hepatitis types in untreated wastewater. The majority of contigs aligned were annotated as viruses. The use of *in vitro* virus concentration methods has been known to improve virus identification in metagenomes in previous studies (Hjelmsø et al., 2017; McCall and Xagoraraki, 2019). Three viral hepatitis types were identified in the untreated wastewater samples: HAV, HEV, and HCV. The consistent detection of HAV on metagenomic samples in parallel with qPCR findings suggest the potential of NGS and metagenomics to be employed as a tool for viral disease surveillance. HAV was identified in all 18 samples with the greatest relative abundance compare to HEV and HCV. To explore the impact

of disease incidence on detection rate in sequenced samples, sequencing can be performed during non-epidemic conditions. This can further act to determine the quiet circulation of viral hepatitis types during periods of no or low clinical activity.

Like HAV, HEV is an enteric virus causing acute hepatitis and is transmitted through contaminated water and food. Albeit, HEV has a low fatality rate in immunocompetent individuals, pregnant women and individuals with suppressed immune systems could face chronic infections and higher death rates (Kamar et al., 2014). Despite, HEV infections being rare in the U.S., metagenomic findings indicate the presence of the virus in 72 % of untreated wastewater with 7 cases reported in the service community in 2017 and 2018 combined. It is possible that asymptomatic infections have occurred in the study community or the virus persist in the environment and deserves further investigation.

Furthermore, HCV, a bloodborne pathogen usually causing chronic hepatitis, was discovered in 2 samples. According to the CDC, approximately 2.4 million people in the U.S. are living with chronic hepatitis C, which is the leading cause of liver disease worldwide (Chen and Morgan, 2006). More than 9,000 probable cases of hepatitis C were reported in the service community during the 2017-2018 sampling year. Despite the high number of disease cases HCV obtained the lowest detection rate. Although, HCV was been detected in feces of chronically infected individuals (Beld et al., 2000), the primary route of transmission, location of viral shedding, duration of infection, and presence of environmental inhibitors can affect the presence of HCV in wastewater. A larger number of samples, collection of untreated sewage from localized areas, and

detection using sensitive molecular assays may help to identify patterns of HCV in wastewater and its potential to be monitored using wastewater surveillance methods.

4.4. WBE for Early Detection of Viral Disease Outbreaks

Several approaches have been implemented for estimating viral presence in the surrounding community. These include using simple mathematical models to predict the incidence of cases given viral concentrations (Hellmér et al., 2014) or predict viral concentrations in wastewater given the number of reported cases and estimated viral excretion rates (Miura et al., 2016). Additionally, models using virus behavior and knowledge of disease transmission rates have demonstrated good accuracy when predicting epidemic patterns of disease (Brouwer et al., 2018). This study has demonstrated the use of WBE for early detection of hepatitis A outbreaks using a simple mechanistic model to account for viral disease patterns. With robust datasets, statistical models can be used to make general inferences about the effects of an outbreak on the behavior of the causative agent in untreated community wastewater.

It is important to note that spatial discrepancy between health data and the WRRF service community poses limitations on the study. Additionally, travelers or those who work within the service area but live and seek medical care outside of the catchment area can lead to misalignments between environmental and clinical data. Frequent localized sampling, daily disease counts from neighboring health clinics, and analysis of stable biomarkers could lead to better insights and improved sensitivity of the method.

5. Conclusion

Generally, molecular methods implemented in this investigation effectively captured HAV loads in wastewater during peak and post peak hepatitis A outbreak conditions. Hepatitis A cases were strongly correlated with viral concentrations in wastewater during peak outbreak conditions when adjusting for disease patterns. Increases in hepatitis A incidence in the surrounding community were revealed in wastewater approximately 7 to 9 days before cases were reported to health care facilities. Moreover, the sensitivity of WBE to capture a rise in disease occurrence depends on the extent of cases present within the community. Despite strong correlations between clinical cases and HAV viral concentrations in wastewater, more frequent and rigorous environmental sampling is needed to fully understand HAV patterns in wastewater under various conditions. Additionally, statistical models used for establishing associations between virus concentrations and disease presence can vary greatly depending on locality. Establishing a baseline for HAV concentrations in wastewater within a given region can help to distinguish between environmental background and outbreak conditions. Such efforts can provide the basis for establishing actionable HAV concentration thresholds in wastewater for public health officials. Lastly, metagenomics detected the presence of three viral hepatitis types in untreated wastewater samples, these were HAV, HEV, and HCV. This demonstrates that molecular and sequencing approaches can work together to identify various human viruses circulating in the community, better forecast disease outbreaks, and facilitate monitoring strategies for disease prevention.

Acknowledgements

This study was funded by National Science Foundation Award #1752773. We thank Anil Gosine (Detroit Water and Sewerage Department) and Michael Jurban (Great Lakes Water Authority) for allowing access to the Detroit wastewater treatment utility and assisting with sampling. We thank the Research Technology Support Facility (Michigan State University) for assisting with sequencing and Professor Shinhan Shiu (Michigan State University) for assisting with bioinformatics analysis. We thank Professor Hui Li (Michigan State University) for assisting with biomarker analysis.

APPENDICES

APPENDIX A2: Supplementary Methods

Population Estimation

The number of people served during each sampling event was estimated according to equation (4), modified from Rico et al., 2017.

$$Pop. served = [b * Q * p]/E$$
(4)

where b is the biomarker concentration (mg/L), Q is the average wastewater flowrate (liters/day), p is the percent of flow in the respective interceptor, and E is the biomarker excretion rate in (mg/person/day). Average flowrate and percent of flow in each interceptor was determined by Detroit Water and Sewage Department personnel, and the Great Lakes Water Authority Capital Improvement Plan (GLWA, 2018). Creatinine excretion rates were calculated according to Ix et al., 2011 and Walser, 1987. Concentrations using the Ix consider both race and sex, while the Walser method considers sex only.

Due to insufficient anthropometric data for Michigan, the United States was considered for excretion calculations. A past study revealed similar weight profiles between Michigan and the United States for adult men and women (Moffatt et al., 1980), thus making the U.S. a feasible substitute for investigating population estimates. Demographic and weight data for the U.S. were extracted from the 2018 census data (U.S. Census Bureau 2018), and CDC's anthropometric reference data for adults 20 years and older (Fryar et al., 2016).

Creatinine concentrations determined based on methods used from Ix et al., 2011 and Walser, 1987 yielded similar excretion rates. Therefore, single-race blacks and whites were averaged for both males and females. Rates for blacks were multiplied by a factor of 0.2, which is the ratio of blacks to whites in the U.S. (U.S. Census Bureau 2018). The average creatinine excretion rate was determined by combining the average rates for males and females with each multiplied by the percent of males and females in the U.S.

APPENDIX B2: Supplementary Table and Figures

Sampling Year (SY)	Sampling Date	Sampling Location (Interceptor)	No. of Replicates		
2015 2010 (5111)		DRI	3		
	11/17/2017	NI-EA	3		
		O-NWI	3		
	12/1/2017	DRI	3		
		NI-EA	3		
		O-NWI	3		
	12/14/2017	DRI	3		
		NI-EA	3		
		O-NWI	3		
2017-2018 (51 1)	1/19/2018	DRI	3		
		NI-EA	3		
		O-NWI	3		
		DRI	3		
	2/2/2018	NI-EA	3		
		O-NWI	3		
	2/16/2018	DRI	3		
		NI-EA	3		
		O-NWI	3		
Total for SY 1			54		
2018-2019 (SY 2)	10/17/2018	NI-EA	3		
		O-NWI	3		
	10/31/2018	DRI	3		
		NI-EA	3		
		O-NWI	2		
	11/28/2018	DRI	3		
		NI-EA	3		
		O-NWI	3		
	12/12/2018	DRI	3		
		NI-EA	3		
		O-NWI	3		
	1/17/2019	DRI	2		
		NI-EA	3		
		O-NWI	3		
	2/7/2019	NI-EA	3		
	2/14/2019	NI-EA	3		
	0/00/2010	DRI	3		
	2/28/2019	NI-EA	3		
		DRI	3		
	3/14/2019	NI-EA	3		

Table B2. 1. Summary of sampling schedule and locations.

Table B2.1 (cont'd)

Total for SY 2

Sampling Year 1	Sampling Dates							
Parameters	Units	17-Nov-17	1-Dec-17	14-Dec-17	19-Jan-18	2-Feb-18	16-Feb-18	
O-NWI								
pH		7.3	7.4	6.6	7.3	7.5	7.5	
Temperature	°C	22.7	16.3	15.2	11.7	11.8	8.6	
Dissolved Oxygen	mg/L	2.0	NR	7.7	12.0	13.3	20.2	
Conductivity	µS/cm	1165.0	805.0	1205.0	889.0	1213.3	1254.3	
Nitrate (NO3-N)	mg/L	1.3	0.8	1.6	1.5	1.7	3.2	
NI-EA								
pH		7.3	7.3	6.7	7.3	7.3	7.3	
Temperature	°C	18.0	18.6	16.1	13.9	12.9	9.4	
Dissolved Oxygen	mg/L	1.7	-	7.4	2.4	2.8	11.9	
Conductivity	µS/cm	987.0	989.0	1260.0	1072.0	1070.0	972.7	
Nitrate (NO3-N)	mg/L	1.0	0.6	0.7	1.0	1.5	1.9	
DRI								
pH		7.3	7.4	6.6	7.1	7.4	7.0	
Temperature	°C	15.3	13.4	12.0	9.3	9.0	8.1	
Dissolved Oxygen	mg/L	2.0	6.3	4.9	21.6	18.8	25.2	
Conductivity	µS/cm	624.0	638.0	1103.0	757.0	992.0	1383.0	
Nitrate (NO3-N)	mg/L	0.7	0.8	0.9	0.6	1.8	2.5	

Table B2. 2. Wastewater physio-chemical characteristics per sampling date and location for sampling year 1.

Sampling Year 2	Sampling Dates									
Parameters	Units	17-Oct-18	31-Oct-18	28-Nov-18	12-Dec-18	17-Jan-19	7-Feb-19	17-Feb-19	28-Feb-19	14-Mar-19
O-NWI										
pH		7.0	6.8	7.3	7.5	7.2	NR	NR	NR	NR
Temperature	°C	18.5	16.0	13.4	13.0	12.4	NR	NR	NR	NR
Dissolved Oxygen	mg/L	NR	NR	18.2	12.3	9.4	NR	NR	NR	NR
Conductivity	μS/cm	1154.0	406.0	1108.0	1210.0	781.2	NR	NR	NR	NR
Nitrate (NO3-N)	mg/L	NR	NR	NR	NR	NR	NR	NR	NR	NR
NI-EA										
pH		7.0	6.8	7.2	7.2	7.0	7.5	7.2	7.1	7.3
Temperature	°C	18.9	16.5	16.4	16.9	14.6	13.2	15.0	12.4	12.4
Dissolved Oxygen	mg/L	NR	NR	2.0	16.0	16.6	2.3	1.1	1.8	0.6
Conductivity	µS/cm	1132.0	464.0	1064.0	1002.0	903.7	1432.7	1573.0	1302.0	987.0
Nitrate (NO3-N)	mg/L	NR	2.2	NR	NR	NR	NR	NR	NR	NR
DRI										
pH		NR	6.8	7.2	7.2	7.1	NR	NR	7.1	7.7
Temperature	°C	NR	15.0	13.2	13.0	11.5	NR	NR	8.7	8.9
Dissolved Oxygen	mg/L	NR	NR	1.7	1.3	8.5	NR	NR	20.4	15.8
Conductivity	µS/cm	NR	355.0	780.0	785.0	996.7	NR	NR	1063.0	842.3
Nitrate (NO3-N)	mg/L	NR	2.8	NR	NR	NR	NR	NR	NR	NR

Table B2. 3. Wastewater physio-chemical characteristics per sampling date and location for sampling year 2.



Figure B2. 1. Virus sampling setup with electropositive NanoCeram cartridge filters.



Figure B2. 2. Boxplots of creatinine concentrations in parts per billion (ppb) per sampling date for sampling year 1 (A) and 2 (B). Boxplot of creatinine concentrations per interceptor (C). Mean concentration is denoted as x, median concentration is denoted with a horizontal line.



Figure B2. 3. Estimated number of people represented in wastewater sampled for each sampling event. Population estimated are based on creatinine concentrations. Error bars represent standard error. Red arrows indicate peaks. Peaks may not be statistically significant. Line breaks indicate missing data.





Figure B2. 4. Average daily precipitation in inches in Wayne, Oakland, and Macomb counties during sampling years 1 (A) and 2 (B). Precipitation represents rainfall and snowmelt. Data extracted from the National Oceanic and Atmospheric Administration's (NOAA's) Global Historical Climatology Network (GHCN) database (Menne et al., 2012).

HAV Concentrations: Histogram



Figure B2. 5. Histogram of hepatitis A virus (HAV) concentrations from sampling year 1 and 2. Dotted lines represent average HAV concentration. Concentrations below the limit of detection have been replaced with one half the detection limit. Concentrations represented here include all data (i.e. outliers).



Figure B2. 6. Spearman's correlation analysis plots for hepatitis A virus concentrations in wastewater and number of confirmed hepatitis A cases for sampling year 1 (A) and 2 (B). Gray shaded region represents 95% confidence interval.



Figure B2. 7. Quantile-quantile plot for evaluating normality of log transformed hepatitis A virus concentrations. Data represented include sampling year 1 and 2 with erroneous outlier points removed according to Cook's distance.



Figure B2. 8. Concentrations of hepatitis A virus per biological replicate (BR) in each interceptor over time for sampling year 1 and 2. Lines representation linear regression lines. NI-EA interceptor renamed to aNI-EA for reordering purposes. Time is indicated as sampling week.



Figure B2. 9. Distribution of hepatitis A virus concentrations per sampling year for each sampling event. Time is indicated as sampling week. Lines produced from linear regression analysis. Shaded regions represent 95% confidence interval.



Figure B2. 10. Comparison of Hepatitis A concentrations and number of cases for sampling year 1 and 2. Line representation linear regression lines.

```
Call:
lm(formula = log(HAV) \sim CASES + PRECP + INTCEP, data = HAV_OR)
Residuals:
             1Q Median
    Min
                             3Q
                                    Мах
-5.1216 -1.2208 0.0857 1.3119
                                3.9049
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.35471
                       0.33127 37.295 < 2e-16 ***
                                 6.791 7.32e-10 ***
             0.13364
                       0.01968
CASES
                       0.40379
                                 2.278
PRECP
             0.91988
                                         0.0248 *
           -1.10733
                       0.42920 -2.580
                                         0.0113 *
INTCEPDRI
                       0.46178 -0.403
INTCEPO-NWI -0.18595
                                         0.6880
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1.898 on 103 degrees of freedom
  (23 observations deleted due to missingness)
Multiple R-squared: 0.3636, Adjusted R-squared: 0.3389
F-statistic: 14.71 on 4 and 103 DF, p-value: 1.536e-09
```

Figure B2. 11. Summary of stepwise multiple linear regression analysis. Final model. Regression analysis performed in R.



Figure B2. 12. Diagnostics plots obtained from the final linear regression model.



Figure B2. 13. Assessment of remaining outliers using Cook's distance (A) and Cook's distance vs. leverage (B) plots. All outliers were considered quality measured concentrations.



Figure B2. 14. Total number of contigs assigned to each virus group during metagenomic analysis. Total number includes all 18 samples for sampling year 1. Further analysis was performed on the Riboviria realm.

REFERENCES

REFERENCES

Andrews, S., 2010. FastQC. Babraham Bioinforma.

- Aw, T.G., Howe, A., Rose, J.B., 2014. Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater. J. Virol. Methods.
- Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67.
- Beld, M., Sentjens, R., Rebers, S., Weel, J., Wertheim-Van Dillen, P., Sol, C., Boom, R., 2000. Detection and quantitation of hepatitis C virus RNA in feces of chronically infected individuals. J. Clin. Microbiol. 38, 3442–3444.
- Bibby, K., Viau, E., Peccia, J., 2011. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. Lett. Appl. Microbiol.
- Bisseux, M., Colombet, J., Mirand, A., Roque-Afonso, A.M., Abravanel, F., Izopet, J.,
 Archimbaud, C., Peigue-Lafeuille, H., Debroas, D., Bailly, J.L., Henquell, C., 2018.
 Monitoring human enteric viruses in wastewater and relevance to infections encountered in the clinical setting: A one-year experiment in central France, 2014 to 2015.
 Eurosurveillance 23, 1–11.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., Schneider, M., 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 31, 365–370.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics.
- Brouwer, A.F., Eisenberg, J.N.S., Pomeroy, C.D., Shulman, L.M., Hindiyeh, M., Manor, Y., Grotto, I., Koopman, J.S., Eisenberg, M.C., 2018. Epidemiology of the silent polio outbreak in Rahat, Israel, based on modeling of environmental surveillance data. Proc. Natl. Acad. Sci. U. S. A. 115, E10625–E10633.
- Burns, M., Valdivia, H., 2008. Modelling the limit of detection in real-time quantitative PCR. Eur. Food Res. Technol. 226, 1513–1524.
- Casto, A.M., Adler, A.L., Makhsous, N., Crawford, K., Qin, X., Kuypers, J.M., Huang, M.L., Zerr, D.M., Greninger, A.L., 2019. Prospective, real-time metagenomic sequencing during norovirus outbreak reveals discrete transmission clusters. Clin. Infect. Dis. 69, 941–948.
- CDC, 2015. Epidemiology and Prevention of Vaccine-Preventable Diseases, 13th ed. Washington D.C. Public Health Foundation.

- CDC, 2020. Widespread person-to-person outbreaks of hepatitis A across the United States [WWW Document]. Viral Hepat. URL https://www.cdc.gov/hepatitis/outbreaks/2017March-HepatitisA.htm (accessed 4.27.20).
- Chen, C., Kostakis, C., Gerber, J.P., Tscharke, B.J., Irvine, R.J., White, J.M., 2014. Towards finding a population biomarker for wastewater epidemiology studies. Sci. Total Environ.
- Chen, S.L., Morgan, T.R., 2006. The natural history of hepatitis C virus (HCV) infection. Int. J. Med. Sci. 3, 47–52.
- Chen, W.C., Chiang, P.H., Liao, Y.H., Huang, L.C., Hsieh, Y.J., Chiu, C.M., Lo, Y.C., Yang, C.H., Yang, J.Y., 2019. Outbreak of hepatitis a virus infection in Taiwan, june 2015 to september 2017. Eurosurveillance 24, 1–10.
- Davis, M.L., 2010. Water and wastewater engineering Design Principles and Practice, Wetpress.
- Farkas, K., Cooper, D.M., McDonald, J.E., Malham, S.K., de Rougemont, A., Jones, D.L., 2018a. Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine receiving waters. Sci. Total Environ. 634, 1174–1183.
- Farkas, K., Marshall, M., Cooper, D., McDonald, J.E., Malham, S.K., Peters, D.E., Maloney, J.D., Jones, D.L., 2018b. Seasonal and diurnal surveillance of treated and untreated wastewater for human enteric viruses. Environ. Sci. Pollut. Res. 25, 33391–33401.
- Fernandez-Cassi, X., Timoneda, N., Martínez-Puchol, S., Rusiñol, M., Rodriguez-Manzano, J., Figuerola, N., Bofill-Mas, S., Abril, J.F., Girones, R., 2018. Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. Sci. Total Environ. 618, 870–880.
- Fiore, A.E., 2004. Hepatitis A Transmitted by Food. Clin. Infect. Dis. 38, 705–715.
- Gharbi-khelifi, H., Sdiri, K., Ferre, V., Harrath, R., Berthome, M., Billaudel, S., Aouni, M., 2007. A 1-year study of the epidemiology of hepatitis A virus in Tunisia. Clin. Microbiol. Infect.
- GLWA, 2018. Capital Improvement Plan 2019-2023.
- Greninger, A.L., Naccache, S.N., Federman, S., Yu, G., Mbala, P., Bres, V., Stryke, D., Bouquet, J., Somasekar, S., Linnen, J.M., Dodd, R., Mulembakani, P., Schneider, B.S., Muyembe-Tamfum, J.J., Stramer, S.L., Chiu, C.Y., 2015. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 7.
- Hellmér, M., Paxéus, N., Magnius, L., Enache, L., Arnholm, B., Johansson, A., Bergström, T., Norder, H., 2014. Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. Appl. Environ. Microbiol. 80, 6771–6781.

- Hjelmsø, M.H., Hellmér, M., Fernandez-Cassi, X., Timoneda, N., Lukjancenko, O., Seidel, M., Elsässer, D., Aarestrup, F.M., Löfström, C., Bofill-Mas, S., Abril, J.F., Girones, R., Schultz, A.C., 2017. Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing. PLoS One 12, 1–17.
- Hofmeister, M.G., Yin, S., Aslam, M. V., Teshale, E.H., Spradling, P.R., 2020. Hepatitis A Hospitalization Costs, United States, 2017. Emerg. Infect. Dis. 26, 1040–1041.
- Huson, D.H., Albrecht, B., Bağci, C., Bessarab, I., Górska, A., Jolic, D., Williams, R.B.H., 2018. MEGAN-LR: New algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. Biol. Direct.
- Jothikumar, N., Cromeans, T.L., Sobsey, M.D., Robertson, B.H., 2005. Development and evaluation of a broadly reactive TaqMan assay for rapid detection of hepatitis A virus. Appl. Environ. Microbiol. 71, 3359–3363.
- Kamar, N., Dalton, H.R., Abravanel, F., Izopet, J., 2014. Hepatitis E virus infection. Clin. Microbiol. Rev. 27, 116–138.
- Kamel, A.H., Ali, M.A., El-Nady, H.G., Aho, S., Pothier, P., Belliot, G., 2010. Evidence of the co-circulation of enteric viruses in sewage and in the population of Greater Cairo. J. Appl. Microbiol.
- Kim, T.D., Unno, H., 1996. The roles of microbes in the removal and inactivation of viruses in a biological wastewater treatment system. In: Water Science and Technology.
- Kokkinos, P., Ziros, P., Meri, D., Filippidou, S., Kolla, S., Galanis, A., Vantarakis, A., 2011. Environmental surveillance. An additional/alternative approach for virological surveillance in Greece? Int. J. Environ. Res. Public Health 8, 1914–1922.
- La Rosa, G., Della Libera, S., Iaconelli, M., Ciccaglione, A.R., Bruni, R., Taffon, S., Equestre, M., Alfonsi, V., Rizzo, C., Tosti, M.E., Chironna, M., Romanò, L., Zanetti, A.R., Muscillo, M., 2014. Surveillance of hepatitis A virus in urban sewages and comparison with cases notified in the course of an outbreak, Italy 2013. BMC Infect. Dis. 14, 1–11.
- Lemon, S.M., Ott, J.J., Van Damme, P., Shouval, D., 2018. Type A viral hepatitis: A summary and update on the molecular virology, epidemiology, pathogenesis and prevention. J. Hepatol. 68, 167–184.
- Manor, Y., Lewis, M., Ram, D., Daudi, N., Mor, O., Savion, M., Kra-Oz, Z., Avni, Y.S., Sheffer, R., Shouval, D., Mendelson, E., 2017. Evidence for hepatitis a virus endemic circulation in Israel despite universal toddler vaccination since 1999 and low clinical incidence in all age groups. J. Infect. Dis. 215, 574–580.
- Matus, M., Duvallet, C., Soule, M.K., Kearney, S.M., Endo, N., Ghaeli, N., Brito, I., Ratti, C., Kujawinski, E.B., Alm, E.J., 2019. 24-hour multi-omics analysis of residential sewage

reflects human activity and informs public health. bioRxiv.

- McCall, C., Xagoraraki, I., 2019. Metagenomic approaches for detecting viral diversity in water environments. J. Environ. Eng. 145.
- MDHHS, 2017. Michigan's communicable disease rules [WWW Document]. URL https://www.michigan.gov/mdhhs/0,5885,7-339-71551_2945_5103_26138-15166--,00.html (accessed 4.30.20).
- MDHHS, 2020a. Michigan Disease Surveillance System background [WWW Document]. URL https://www.michigan.gov/mdhhs/0,5885,7-339-71550_5104_31274-96814--,00.html (accessed 4.28.20).
- MDHHS, 2020b. Michigan hepatitis A outbreak: Update [WWW Document]. URL https://www.michigan.gov/mdhhs/0,5885,7-339-71550_2955_2976_82305_82310-447907--,00.html (accessed 4.27.20).
- Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E., Houston, T.G., 2012. An overview of the global historical climatology network-daily database. J. Atmos. Ocean. Technol. 29, 897– 910.
- Miura, T., Lhomme, S., Le Saux, J.C., Le Mehaute, P., Guillois, Y., Couturier, E., Izopet, J., Abranavel, F., Le Guyader, F.S., 2016. Detection of Hepatitis E Virus in Sewage After an Outbreak on a French Island. Food Environ. Virol.
- Munir, M., Wong, K., Xagoraraki, I., 2011. Release of antibiotic resistant bacteria and genes in the effluent and biosolids of five wastewater utilities in Michigan. Water Res.
- Ng, T.F.F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B.S., Wommack, K.E., Delwart, E., 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J. Virol. 86, 12161–12175.
- O'Brien, E., Nakyazze, J., Wu, H., Kiwanuka, N., Cunningham, W., Kaneene, J.B., Xagoraraki, I., 2017. Viral diversity and abundance in polluted waters in Kampala, Uganda. Water Res.
- O'Brien, E., Xagoraraki, I., 2019. A water-focused one-health approach for early detection and prevention of viral outbreaks. One Heal.
- Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics.
- Pinon, A., Vialette, M., 2019. Survival of viruses in water. Intervirology 61, 214–222.
- R Core Team, 2019. R: A language and environment for statistical computing. Vienna, Austria.

Sidhu, J.P.S., Sena, K., Hodgers, L., Palmer, A., Toze, S., 2018. Comparative enteric viruses and
coliphage removal during wastewater treatment processes in a sub-tropical environment. Sci. Total Environ.

- Simmons, F.J., Xagoraraki, I., 2011. Release of infectious human enteric viruses by full-scale wastewater utilities. Water Res. 45, 3590–3598.
- Snyder, M.R., McGinty, M.D., Shearer, M.P., Meyer, D., Hurtado, C., Nuzzo, J.B., 2019. Outbreaks of hepatitis A in US communities, 2017-2018: Firsthand experiences and operational lessons from public health responses. Am. J. Public Health.
- Spierto, F.W., Hannon, W.H., Gunter, E.W., Smith, S.J., 1997. Stability of urine creatinine. Clin. Chim. Acta.
- Thai, P.K., O'Brien, J., Jiang, G., Gernjak, W., Yuan, Z., Eaglesham, G., Mueller, J.F., 2014. Degradability of creatinine under sewer conditions affects its potential to be used as biomarker in sewage epidemiology. Water Res. 55, 272–279.
- Tolouei, S., Autixier, L., Taghipour, M., Burnet, J.B., Bonsteel, J., Duy, S.V., Sauvé, S., Prévost, M., Dorner, S., 2019. Precipitation effects on parasite, indicator bacteria, and wastewater micropollutant loads from a water resource recovery facility influent and effluent. J. Water Health 17, 701–716.
- U.S. EPA, 2001. Manual of Methods for Virology (Chapter 14).
- Vajargah, K.F., Nikbakht, M., 2015. Application REMLModel and determining cut off of ICC by multi-level model based on Markov Chains and simulation in health. Indian J. Fundam. Appl. Life Sci. 5, 1432–1448.
- Varughese, E.A., Brinkman, N.E., Anneken, E.M., Cashdollar, J.L., Fout, G.S., Furlong, E.T., Kolpin, D.W., Glassmeyer, S.T., Keely, S.P., 2018. Estimating virus occurrence using Bayesian modeling in multiple drinking water systems of the United States. Sci. Total Environ. 619–620, 1330–1339.
- Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S Fourth edition by, World.
- Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., Mardis, E.R., Hickenbotham, M., Magrini, V., Eldred, J., Latreille, J.P., Wilson, R.K., Ganem, D., DeRisi, J.L., 2003. Viral discovery and sequence recovery using DNA microarrays. PLoS Biol.
- WHO, 2017. Global hepatitis report, 2017.
- Xagoraraki, I., O'Brien, E., 2020. Wastewater-based epidemiology for early detection of viral outbreaks.

Yanez, L.A., Lucero, N.S., Barril, P.A., Díaz, M. del P., Tenaglia, M.M., Spinsanti, L.I., Nates,

S. V., Isa, M.B., Ré, V.E., 2014. Evidence of Hepatitis A virus circulation in central Argentina: Seroprevalence and environmental surveillance. J. Clin. Virol. 59, 38–43.

- Zhang, X.S., Iacono, G. Lo, 2018. Estimating human-to-human transmissibility of hepatitis A virus in an outbreak at an elementary school in China, 2011. PLoS One 13, 1–14.
- Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. Methods Ecol. Evol.

CHAPTER 3: IDENTIFICATION OF MULTIPLE POTENTIAL VIRAL DISEASES IN A LARGE URBAN CENTER USING WASTEWATER SURVEILLANCE

Published in Water Research:

McCall, C., Wu, H., Miyani, B., Xagoraraki, I., 2020. Identification of multiple potential viral diseases in a large urban center using wastewater surveillance. Water Res. 184.

Abstract

Viruses are linked to a multitude of human illnesses and can disseminate widely in urbanized environments causing global adverse impacts on communities and healthcare infrastructures. Wastewater-based epidemiology was employed using metagenomics and quantitative polymerase chain reaction (qPCR) assays to identify enteric and non-enteric viruses collected from a large urban area for potential public health monitoring and outbreak analysis. Untreated wastewater samples were collected from November 2017 to February 2018 (n=54) to evaluate the diversity of human viral pathogens in collected samples. Viruses were classified into virus types based on primary transmission routes and reviewed against viral associated diseases reported in the catchment area. Metagenomics detected the presence of viral pathogens that cause clinically significant diseases reported within the study area during the sampling year. Detected viruses belong to the Adenoviridae, Astroviridae, Caliciviridae, Coronaviridae, Flaviviridae, Hepeviridae, Herpesviridae, Matonaviridae, Papillomaviridae, Parvoviridae, Picornaviridae, Poxviridae, Retroviridae, and Togaviridae families. Furthermore, concentrations of adenovirus, norovirus GII, sapovirus, hepatitis A virus, human herpesvirus 6, and human herpesvirus 8 were measured in wastewater samples and compared to metagenomic findings to confirm detected viral

genus. Hepatitis A virus obtained the greatest average viral load (1.86x10⁷ genome copies/L) in wastewater samples compared to other viruses quantified using qPCR with a 100% detection rate in metagenomic samples. Findings obtained from this study aid in evaluating the utility of wastewater-based epidemiology for identification and routine monitoring of various viruses in large communities. This methodology has the potential to improve public health responses to large scale outbreaks and viral pandemics.

1. Introduction

Viruses are linked to a host of illness related to respiratory infections, diarrheal illness, autoimmune diseases, meningitis, hepatitis, cancer, viral hemorrhagic fevers and others. Infections often disseminate quickly in urbanized regions due to densely populated areas and could reach thousands of inhabitants before health care facilities are notified. Since viruses do not replicate outside of a host and can remain stable in the environment for significant periods of time, wastewater-based epidemiology (WBE) can be used to capture a near real-time picture of the viral disease burden within a community. Viruses can enter waste streams through multiple routes including stool, urine, skin, saliva, and blood, thus wastewater has the potential to assess the burden of a variety of viruses.

It is well known that confirmed enteric viruses, such as rotaviruses, adenoviruses, enteroviruses, hepatitis A and E viruses, caliciviruses, and others can be detected in wastewater. While it is logical to investigate the applicability of enteric viruses to WBE, it is also important to demonstrate the potential for other viruses to fit into this methodology. Indeed, it has been shown that multiple, non-enteric viruses such as coronaviruses, herpesviruses, influenza, zika, West Nile, yellow fever,

dengue and others have been detected in stool and urine samples or wastewater (Barzon et al., 2013; Gourinat et al., 2015; Gundy et al., 2009; Heijnen and Medema, 2011; Hirayama et al., 2012; Hirose et al., 2016; O'Brien and Xagoraraki, 2019; Poloni et al., 2010; Tonry et al., 2005; Xagoraraki and O'Brien, 2020). These observations confirm that the concept of wastewater-based-epidemiology can be applied to a wide range of viruses beyond the confirmed enteric viruses.

Enteric viruses are commonly discovered in untreated wastewater (Farkas et al., 2018; Ng et al., 2012; Victoria et al., 2014) and recent studies have confirmed corresponding disease prevalence in the surrounding community (Bisseux et al., 2018). Particularly, picornaviruses constitute an important group of enteric viruses that cause a host of illnesses including diseases of the central nervous systems, respiratory tract, liver, and gastrointestinal tract. Routine monitoring of picornaviruses in wastewater can provide insight into the transmission of clinically important diseases, prevent widespread outbreaks, and reduce deaths linked to such viruses.

Molecular and sequencing approaches provide qualitative and quantitative insights into wastewater environments (McCall and Xagoraraki, 2019). Metagenomics allows for the screening of a large panel of viruses in environmental systems that would otherwise prove time consuming using traditional laboratory techniques. Metagenomics has identified co-infecting organisms during outbreak conditions (Li et al., 2019), novel pathogens (Cantalupo et al., 2011), and viral compositions in complex matrices (Fancello et al., 2013; Miranda et al., 2016; O'Brien et al., 2017). Despite its breakthroughs, the sensitivity of metagenomics to identify vial pathogens is confounded by the presence of bacteria, sequencing limitations, and errors imposed by sequence analysis and alignment tools. For these reasons sensitive PCR techniques are commonly used to corroborate results obtained from metagenomic analysis.

Here, metagenomic analysis was utilized to assess the diversity of human viral pathogens in untreated wastewater collected from a large urban center over the course of four months. Detected viral pathogens were further classified according to virus type and compared with health data, with an emphasis on picornaviruses, from the associated community to evaluate the application of wastewater-based epidemiology for identification of endemic disease and potential upcoming viral outbreaks. Quantitative PCR and RT-qPCR assays were performed on select viruses commonly present in sewage to corroborate results obtained through metagenomic approaches.

2. Methods

2.1. Study Area and Wastewater Sample Collection

Wastewater surveillance was conducted at the Water Resource Recovery Facility (WRRF) located in Detroit, Michigan. The Detroit WRRF is the largest single site wastewater treatment plant in the U.S. and treats wastewater from an estimated 3 million inhabitants with an average daily flow of 650 MGD (GLWA, 2018). It services the three largest counties, by population, in Michigan. These are Wayne, Oakland, and Macomb counties (Jones et al., 2015). The WRRF receives wastewater from its service municipalities via three main interceptors: North Interceptor-East Arm (NI-EA), Detroit River Interceptor (DRI), and Oakwood-Northwest-Wayne County Interceptor (O-NWI). These interceptors are large sewers that collect and transports wastewater from smaller sewers to the WRRF.

Untreated wastewater samples were collected at the WRRF from sampling points located at each

of the three interceptors approximately bi-weekly between November 2017 and February 2018 (n=54). Viruses were isolated from untreated wastewater using electropositive NanoCeram column filters following the EPA's virus adsorption-elution protocol (U.S. EPA, 2001). Samples were collected in triplicates for each interceptor per sampling date where wastewater was passed through a column filter until fouling occurred. Average filtered sample volumes range between 24-44 liters per interceptor. Each interceptor was sampled with its own filter house, tubing, and vacuum pump to minimize cross contamination. Virus filters were immediately stored on ice and transported to the Environmental Virology Laboratory at Michigan State University (MSU) and stored in -20°C until further processing.

2.2. Sample Processing and Virus Isolation

Following wastewater sampling, NanoCeram cartridge filters were eluted within 24 h with 1.5% w/v beef extract (0.05 M glycine, pH 9.5) according to the EPA's protocol (U.S. EPA, 2001). In short, filters were eluted with 1 L of beef extract for a total of 2 min. The pH of the solution was adjusted to 3.5 ± 0.1 and flocculated for 30 min before centrifugation at 2500g for 15 min at 4°C. Supernatant was discarded and pellets were resuspended in 30 mL of 0.15 M sodium phosphate (pH 9.0-9.5) followed by a second round of centrifugation carried out at 7000g for 10 min at 4°C. The supernatant was neutralized (pH ~7.25) and subjected to filtration using to 0.45µm and 0.22µm syringe filters to eliminate bacterial contamination. Extraction of nucleic acid was performed on 140 µL of purified virus concentrate using the QIAamp Viral RNA Mini Kit (Qiagen) following the manufacturer's protocol and eluted in 80 µL of elution buffer. Nucleic acid was stored at -80°C until further processing.

2.3. Metagenomic Analysis

2.3.1. Sampling Processing and Random amplification

To explore human virus diversity between sampling locations and dates, purified nucleic acid from each biological replicate was pooled together for a total of 18 samples. These samples represent genetic material from all three interceptors during each of the six sampling dates. Nucleic acid from each sample was reverse transcribed and subjected to random amplification as previously described (Wang et al., 2003) to evaluate both RNA and DNA viruses.

2.3.2. Next Generation Sequence Processing

Eighteen samples of viral cDNA were sent to the Research Technology Support Facility Genomics Core at Michigan State University for whole-genome shotgun sequencing (WGS). The Illumina TruSeq Nano DNA Library Preparation Kit was used for all cDNA samples. Library preparation was performed on a Perkin Elmer Sciclone G3 robot according to the manufacturer's recommendations. This was followed by sequencing on an Illumina HiSeq4000 platform generating 150 bp paired-end reads.

2.3.3. Sequence Analysis and Taxonomic Annotation

Sequencing reads generated from WGS were processed on a Unix system through the MSU High Performance Computing Center (HPCC). Raw sequences were analyzed for quality using FastQC, a quality control tool for sequencing data (Andrews, 2010). Sequencing adapters and reads with an average quality score below 20 were removed using Trimmomatic (Bolger et al., 2014). Trimmed reads were assembled with IDBA-UD, a short-read de novo sequence aligner for metagenomic data. Reads were assembled into contigs using an iterative k-mer approach with k-mer sizes ranging between 40 and 120 in increments of 10. The remaining parameters were run at default conditions.

Human virus genomes are relatively small and less abundant in wastewater and therefore may be masked by more dominant bacteria or plant virus genomes. It can also be a challenge for reference databases to maintain updated sequence information for viruses possessing high mutation rates, particularly RNA viruses, leaving room for false negatives and a large percentage of unaffiliated contigs (McCall and Xagoraraki, 2019). To compensate for some of these limitations, an optimized multi-alignment approach was used to improve alignment and annotation of human viral contigs. First, contigs were aligned against the Viral RefSeq database using tBLASTx with an E-value of 10⁻³. This approach has been known to increase human viral discovery in metagenomic datasets (Bibby et al., 2011). Aligned contigs were assigned to the lowest common ancestor (LCA) according to the NCBI's taxonomy with MEGAN (v. 6.15.0). The top 10 percent of BLAST alignments with a minimum bit score of 50 and contig coverage of at least 80% were considered in taxonomic analysis. The remaining parameters were run at default conditions. Reads assigned to virus families containing known human pathogens were extracted and aligned with BLASTx with an E-value of 10⁻⁵ against a custom human virus database containing 5,979 human viral proteins in Swiss-Prot database (Boeckmann et al., 2003). These sequences represented all human viral proteins in the Swiss-Prot database at the time of retrieval (September 2019). Target specific databases can reduce ambiguity and improve pathogens discovery. Furthermore, protein searches

are more effective at capturing remote homology as compared to nucleotide searches (Breitwieser et al., 2018). These optimized detection approaches are important for human viral pathogens, which are often difficult to detect in environmental samples due to their low abundance, small genomes, and high mutation rates. To further increase the potential for pathogen discovery and minimize false negatives, contigs assigned to the virus root were also aligned against the Swiss-Prot database. This method was more effective at capturing select viruses confirmed through qPCR and RT-qPCR compared to contigs solely extracted from human viral families (data not shown). Figure 3.1 displays the metagenomics workflow for human virus identification in wastewater samples.



Figure 3. 1. Metagenomic workflow for human virus identification in wastewater samples.

2.4. Quantification of Select Viruses

Quantitative PCR or RT-qPCR was performed on six viruses, namely, sapovirus (SaV), norovirus (NoV) GII, human adenovirus (HAdV) 40 and 41, hepatitis A virus (HAV), and human herpesvirus 6 (HHV-6) and 8 (HHV-8).

2.4.1. Preparation of Standard Curves

HAV and HAdV were obtained from ATCC for preparation of standard controls. Nucleic acid was extracted from each virus isolate as detailed in the previous section and transformed into One Shot TOP10 chemically competent *Escherichia coli* cells using the TOPO Cloning kit (Invitrogen) following the manufacturer's protocol. Plasmid DNA containing cloned HAV and HAdV was extracted and quantified according to a previous method (Munir et al., 2011). The protocol detailed in step two of the subsequent section was utilized to prepare a standard curve with 10-fold serial dilutions of positive HAV and HAdV controls.

Quantitative synthetic NoV GII RNA, SaV RNA, HHV-6 DNA, and HHV-8 DNA was obtained from ATCC. RNA or DNA was diluted 10-fold and analyzed as described in the following section. Standard curves for HAV, HAdV, NoV GII, SaV, HHV-6, and HHV-8 obtained R-squared values of > 99% and slopes of -3.61, -3.66, -3.82, -3.35, -3.88, and -3.59, respectively. The limit of detection (LOD) for each virus was determined by the lowest point on the standard curve or by the lowest dilution with a 95% positive detection rate in at least 10 replicates (Burns and Valdivia, 2008). HAV, HAdV, NoV, SaV, and HHVs obtained a LOD of 10^1 genome copies/ul.

2.4.2. qPCR and RT-qPCR

Quantitative PCR or RT-qPCR assays were used to establish the concentration of HAV, HAdV, NoV GII, SaV, HHV-6, and HHV-8 in wastewater samples. All assays were performed in triplicates on a Mastercycler ep realplex2 (Eppendorf) in 96-well optical plates. Amplification of cDNA was mediated using Lightcycler 480 Probes Master (Roche) at a concentration of $1 \times$ in all reactions. Sterile nuclease free water was used to meet volume requirements in all reactions. Primers and probes used are shown in Table A3.1.

HAV and SaV was quantified using a two-step RT-qPCR based on a previously described methods (Jothikumar et al., 2005; Oka et al., 2006). Briefly, viral RNA was reverse transcribed using iScript RT-qPCR Supermix (Bio-Rad) according to the manufacturer's protocol. For HAV, 5 µL of cDNA, negative control, or positive control was transferred to a 15 µl reaction mix containing HAV primers and TaqMan probe. Reactions were performed with the following conditions: 95°C for 15 min, followed by 45 cycles of 95°C for 15 s, 55°C for 20 s, and 72°C for 15 s. SaV quantification was carried out in a 25 uL reaction containing each primer and probe. Reactions were performed with the followed by 45 cycles of 94°C for 15 s, 62°C for 1 min, and 72°C for 15 s.

Norovirus GII was quantified using a one-step RT-qPCR as previously described (Le Guyader et al., 2009). In short, the RT-qPCR was carried out in a 25 μ L reaction mixture containing primers and probe, 2 μ L of iScript RT-qPCR Supermix, and 5 μ L of viral RNA, negative control, or positive control. Reactions were performed with the following conditions: reverse transcription at

25°C for 5 min, 46°C for 20 min, and 95°C for 1 min, followed by 45 cycles of 95°C for 15 s, 60°C for 1 min, 65°C for 1 min.

DNA viruses HAdV, HHV-6, and HHV-8 were quantified according to previously established methods (Gautheret-Dejean et al., 2002; Lallemand et al., 2000; Xagoraraki et al., 2007). HAdV and HHVs were qualified in 20 uL reactions containing 5 uL of DNA or standard control. Denaturation was carried out at 95 °C for 15 min for all DNA viruses followed by 45 cycles at 95°C for 15 s, 60°C for 30 s, and 72°C for 10 s for HAdV; 45 cycles of 95 °C for 15 s and 60 °C for 1 min for HHV-6; and 45 cycles of 95 °C for 15 s and 65 °C for 1 min for HHV-8.

2.5. Health Data Collection

Disease data for all reportable viral diseases for each service county was obtained from the Michigan Department of Health and Human Services (MDHHS). Probable and confirmed case counts were extracted from the Michigan Disease and Surveillance System (MDSS) weekly surveillance reports (WSR). The MDSS is a communicable disease reporting system used to facilitate coordination and sharing of disease surveillance data among multiple shareholders including healthcare providers and medical laboratories (MDHHS, 2020a). Each weekly surveillance report accounts for disease cases reported from Sunday-Saturday of the corresponding week. It is important to note that the WSR uses gastrointestinal illness (GI) and influenza-like illness (ILI) to represent any disease displaying symptoms of this nature. The etiological agent of the disease is unspecified, but could be of viral, bacterial, or parasitic origin. ILI was defined according to the U.S. influenza surveillance system (CDC, 2019a). GI is defined as symptoms related to diarrhea and/or vomiting (MDHHS, 2018a). Diseases reported in WSRs were considered

if is the primary disease of viruses detected in metagenomic samples and if there was at least one case of that disease reported within Macomb, Oakland, or Wayne Counties during the sampling year (2017 - 2018).

2.6. Statistical and Cluster Analysis

Cluster analysis was performed using the Bray-Curtis dissimilarity index in MEGAN on metagenomic samples at the family level within the Swiss-Prot taxonomic analysis to determine similarity between samples. A one-way analysis of variance (ANOVA) and Tukey's HSD post hoc tests were used to investigate significance between mean concentrations of select viruses in wastewater samples. All statistical analyses were performed in R (R Core Team, 2019).

3. Results

Illumina sequencing was performed on 18 untreated wastewater samples collected from Detroit's WRRF from November 2017 to February 2018. A total of 624.4 million reads were subject to quality trimming resulting in 595.2 million reads. Reads were assembled and aligned using tBLASTx against the Viral RefSeq database. The proportion of contigs assigned to viral taxonomic groups range between 72-83% (Table A3.2). As expected, viruses infecting prokaryotes constituted the greatest proportion of viral reads (Table 3.1).

Table 3. 1. Total number of contigs per virus family with associated host for 18 sequenced samples.

Family	Host	Total No. of Contigs
Siphoviridae	Prokaryotes	627615
Myoviridae	Prokaryotes	626034

Table 3.1 (cont'd)

Podoviridae	Prokaryotes	360840
Microviridae	Prokaryotes	62058
Herelleviridae	Prokaryotes	22218
unclassified bacterial viruses	Prokaryotes	18327
unclassified Caudovirales	Prokaryotes	13681
Inoviridae	Prokaryotes	9085
Ackermannviridae	Prokaryotes	6670
Leviviridae	Prokaryotes	2362
unclassified archaeal viruses	Prokaryotes	87
Lipothrixviridae	Prokaryotes	18
Bicaudaviridae	Prokaryotes	28
Iridoviridae	Animal Invertebrates	3874
Baculoviridae	Animal Invertebrates	849
Ascoviridae	Animal Invertebrate	299
Polydnaviridae	Animals Invertebrates	29
Dicistroviridae	Animals Invertebrates	28
Nudiviridae	Animals Invertebrates	17
Herpesviridae	Animal Vertebrates (includes Humans)	6742
Poxviridae	Animal Vertebrates (includes Humans)	5854
Parvoviridae	Animal Vertebrates (includes Humans)	875
Retroviridae	Animal Vertebrates (includes Humans)	50
Circoviridae	Animal Vertebrates	1297
Alloherpesviridae	Animal Vertebrates	388
Phycodnaviridae	Plants	56073
Virgaviridae	Plants	168
Potyviridae	Plants	108
Caulimoviridae	Plants	27
Mimiviridae	Protists	26077
Marseilleviridae	Protists	7009
Lavidaviridae	Other	1623
unclassified DNA viruses	Other	35010
unclassified Riboviria	Other	9353
unclassified viruses	Other	4672
Genomoviridae	Other	362

3.1. Classification of Human Viral Pathogens in Wastewater

Putative human viral contigs were extracted via MEGAN taxonomic bins and aligned against all human viral proteins in the Swiss-Prot database using BLASTx. An average of approximately 0.18% (0.05-0.78%) of viral affiliated contigs were aligned to human viral proteins. Fourteen different human viral families were identified with the greatest number of contigs largely assigned to *Poxviridae*, *Herpesviridae*, and *Picornaviridae* (Figure 3.2). Of the fourteen families, nine were classified at ssRNA viruses and the remining five as DNA viruses. Figure 3.3 shows the proportion of ssRNA and DNA viral families in each sample. Comparable relative abundances of ssRNA and DNA viruses were observed during the 14 Dec, 19 Jan, and 2 Feb sampling dates with DNA viruses dominating sequenced samples during the remaining three sampling dates.



Figure 3. 2. ssRNA (a) and DNA (b) virus diversity and relative abundance in wastewater samples.

Bray-Curtis dissimilarity analysis was used to determine the similarity between samples at the family taxonomic level after alignment against human viral protein sequences. According to the Bray-Curtis analysis, there were more similarities within sampling dates rather than sampling locations with samples collected during three consecutive sampling dates (14-Dec.,19-Jan., and 2-Feb.) clustering together (Figure 3.3). *Poxviridae*, *Parvoviridae*, and *Herpesviridae* families were the most influential viral families when discriminating between samples (data not shown).



Figure 3. 3. Principal component analysis (PCoA) of human viral pathogen presence. PCoA was produced in MEGAN at the family level using Bray-Curtis dissimilarity index.

Of the 14 human viral families identified, contigs were assigned to 26 human virus genera with DNA viruses, orthopoxivirus, simplex viruses, and lymphocrytovirus obtaining the greatest number of hits. Alphavirus, a ssRNA virus containing vector-borne viruses was the fourth most abundant genus (Figure 3.4a).



Figure 3. 4. (a) Heatmap of human genus virus diversity and normalized abundance in each sample. White cells indicate absence of associated virus in related sample. Virus genera are in descending order according to abundance. Heatmap was produced in R. (b) Proportion of human virus types detected in wastewater samples.

The most frequently detected virus type was enteric and respiratory, followed by other, bloodborne, and vector-borne (Figure 3.4b, Table 3.2). Four of ten enteric viruses detected belong to the *Picornaviridae* family, namely, hepatovirus, enterovirus, parechovirus, and cardiovirus. Figure 3.5 illustrates the temporal and spatial relative abundance of picornaviruses during the study period. Proteins associated with hepatovirus, mamastrovius, and enterovirus obtained the greatest number of assigned contigs within the enteric virus group (Figure 3.4a). Within the mamastrovirus genus, classic human astroviruses (HAstV) were detected with >90% homology to the reference sequence in all positive samples (15/18). Enteroviruses, containing respiratory and enteric pathogens, revealed human coxsackievirus and poliovirus species detected in positive samples with > 80% identity to the reference gene.

114

Table 3. 2. Summary of human viral pathogens detected in wastewater and their associated disease reported in the Michigan Disease Surveillance System (MDSS) Weekly Surveillance Reports (WSR). Associated disease is considered if at least one case was reported during the sampling year (2017-2018).

Mea	surements	MDSS R	leports ^a	Other		
Virus Family	Virus Genus	Specific Primary Reported Disease	Non-specific Reported Illness	Primary Virus Type (Transmission Route)	Oncogenic?	References
Adenoviridae	Mastadenovirus		GI; IFI; Encephalitis, Primary	Enteric; Respiratory		Ghebremedhin, 2014
Astroviridae	Mamastrovirus		GI	Enteric		Bosch et al., 2014
Caliciviridae	Norovirus	Norovirus	GI	Enteric		Glass et al., 2009
	Sapovirus		GI	Enteric		Oka et al., 2015
Coronaviridae	Betacoronavirus	Novel Coronavirus	ILI; GI	Respiratory		(Chan et al., 2015; Kuiken et al., 2003; Lai et al., 2020)
Flaviviridae	Hepacivirus	Hepatitis C		Bloodborne	Y	Chen and Morgan, 2006
Hepeviridae	Orthohepevirus	Hepatitis E	Guillain-Barre Syndrome	Enteric		Kamar et al., 2014; Van Den Berg et al., 2014
Herpesviridae	Lymphocryptovirus		Encephalitis, Primary; Guillain- Barre Syndrome; ILI	Bloodborne; Other	Y	Tsao et al., 2015; Van Den Berg et al., 2014
	Roseolovirus		Encephalitis, Primary; GI; ILI	Other; Respiratory		De Bolle et al., 2005; Hall et al., 1994
	Simplexvirus		Encephalitis, Primary; ILI; Meningitis - Aseptic	Other		Whitley, 2011; Widener and Whitley, 2014
	Varicellovirus	Chickenpox; Shingles; VZ infection unspecified; Encephalitis, Post Chickenpox		Respiratory		Arvin, 1996

Table 3. 2 (cont'd)

Matonaviridae	Rubivirus	Rubella	Guillain-Barre Syndrome; Encephalitis, Primary; ILI	Respiratory		Mawson and Croft, 2019
Papillomaviridae	Alphapapillomavirus			Other	Y	Doorbar et al., 2015
Parvoviridae	Bocaparvovirus		ILI; GI	Enteric; Respiratory		Qiu et al., 2017
	Erythroparvovirus		ILI	Bloodborne; Respiratory		Qiu et al., 2017
Picornaviridae	Cardiovirus		ILI; GI; Encephalitis, Primary	Enteric		Tan et al., 2017
	Enterovirus		Encephalitis, Primary; ILI; GI; Acute Flaccid Myelitis (AFM); Meningitis - Aseptic	Enteric; Respiratory		Wells and Coyne, 2019
	Hepatovirus	Hepatitis A	Guillain-Barré syndrome	Enteric		Lemon et al., 2018
	Parechovirus		ILI; GI; Meningitis - Aseptic	Enteric; Respiratory		de Crom et al., 2016
Poxviridae	Orthopoxvirus		ILI	Other; Respiratory		Buller and Palumbo, 1991; Haller et al., 2014
	Parapoxvirus			Other		Buller and Palumbo, 1991; Fox et al., 2002
Retroviridae	Deltaretrovirus			Bloodborne	Y	Gonçalves et al., 2010; Ishitsuka and Tamura, 2014
	Gammaretrovirus					Denner, 2010
	Lentivirus	HIV		Bloodborne		del Rio, 2017

Table 3. 2 (cont'd)

		Eastern equine		Armstrong and
Togaviridae	Alphavirus	encephalitis;	Vector-borne	Andreadis, 2013;
	-	Chikungunya		Weaver et al., 2018

a: Influenza-like illness (ILI) is defined according to the U.S. influenza surveillance systems (CDC, 2019a). Gastrointestinal illness (GI) is defined as symptoms related to diarrhea and/or vomiting (MDHHS, 2018a). ILI and GI were reported if the virus's primary clinical manifestation is related to that condition. Disease information is not documented for viruses related to non-reportable diseases.



Figure 3. 5. Picornaviruses detected in wastewater samples.

Orthopoxivirus constituted the greatest number of hits within the respiratory group with vaccinia virus being the most frequently identified species. Additionally, metagenomic analysis detected the presence of respiratory pathogen betacoronavirus in 8 of the 18 collected samples with human coronavirus HKU1 being the primary species detected with > 60% percent identity (Figure 3.4a). Other enteric and respiratory viruses detected include mastadenovirus, sapovirus, norovirus, orthohepevirus, varicellovirus, rubivirus, bocaparvovirus, roseolovirus, and erythroparvovirus (Table 3.2).

Viruses solely grouped in the "Other" category consists of viruses transmitted via skin, saliva, or other bodily fluids such as simplex virus, parapoxvirus, and alphapapillomavirus. Furthermore, alphavirus was the only vector-borne virus discovered in metagenomic samples with a detection rate of 89%. Important bloodborne viruses hepacivirus, lentivirus, deltaretrovirus, and lymphocryptovirus were also discovered in metagenomic samples (Figure 3.4, Table 3.2).

3.2. Comparison of Human Viruses and Clinical Data

Weekly surveillance reports from the MDHHS MDSS were utilized to evaluate the potential association between the 26 human viruses detected in metagenomic samples and the presence of primary associated diseases within the surrounding community during the study period. Mastadenovirus, mamastrovirus, sapovirus, bocaparvovirus, cardiovirus, enterovirus, and parechovirus were linked to solely non-specific diseases including flu-like, gastrointestinal illnesses, aseptic meningitis, and acute flaccid myelitis. Norovirus, betacoronavirus, and rubivirus were related to at least one non-specific illnesses along with virus-specific diseases norovirus, novel coronavirus, and rubella, respectively. Virus-specific diseases for hepacivirus, orthohepevirus, varicellovirus, hepatovirus, lentivirus, and alphavirus were present in WSRs during the sampling year. During the time of data collection, mandatory reporting was not required for diseases related to several viruses including parapoxvirus, deltaretrovirus, gammaretrovirus, and alphapapillomavirus (Table 3.2).

3.3. Quantitative Screening for Select Human Viral Pathogens

To confirm results obtained from metagenomic analysis, select viruses were quantified in wastewater samples. Namely, SaV, NoV GII, HAdV 40/41, HAV, HHV-6, and HHV-8. HAV, NoV, HAdV, and HHV-8 were quantified in all samples and obtained positive detection rate of 100%, 50%, 22%, and 0% in metagenomic samples, respectively. SaV and HHV-6 were quantified in 94% and 39% of the 18 samples considered with a 28% and 83% detection in sequenced samples. All select viruses except HHV-6 were detected during each sampling date using qPCR

or RT-qPCR (Figure 3.6). There were significant differences in average concentrations for some viruses where HAV > HAdV > NoV (p < 0.0001). Mean SaV concentrations were significantly less than HAV (p < 0.0001) and greater than NoV (p < 0.0001). There was no significant difference between SaV and HAdV (p > 0.05). Concentrations of HHV-6 and HHV-8 were significantly lower than HAV, HAdV, SaV, and NoV (p < 0.0001). There was no significant difference between mean concentrations of HHV-6 and HHV-8 (p > 0.05) in collected samples.



Figure 3. 6. Boxplot for select viral concentrations per sampling date.

4. Discussion

A WBE study for viral diseases was carried out on wastewater samples collected from a large urban municipal wastewater treatment facility. Samples were subjected to metagenomic analysis and qPCR/RT-qPCR assays to identify human viral pathogens circulating within the community. More than half of all aligned contigs were assigned to viral taxonomic groups. The use of virus enrichment techniques has been known to decrease the presence of non-targeted organisms, such as bacteria and improve virus detection within metagenomes (McCall and Xagoraraki, 2019).

Of the virus affiliated contigs, less than 1% were identified as putative human viral pathogens. Due to their relatively small genomes and low abundance in water reservoirs, human viral reads are known to constitute a small portion of metagenomic datasets (McCall and Xagoraraki, 2019). To compensate for this, a second stage of alignment was carried out on contigs identified as potential human viruses against a custom dataset of human viral proteins. Protein-based alignments are effective at detecting remote homology and therefore allows for the discovery of rapidly evolving viral pathogens (Breitwieser et al., 2018). BLASTx alignment facilitated the taxonomic classification of several different virus types including enteric, respiratory, bloodborne and vector-borne viruses.

4.1. Classification of Human Viruses in Wastewater and Clinical Data Comparison

Enteric and respiratory pathogens were the most frequently detected viruses in sewage samples. Enteric viruses are viruses that mainly infect the intestinal tract and can be transmitted via the fecal-to-oral route. Respiratory viruses generally replicate in the respiratory tract and spread via respiratory secretions. *Picornaviridae* obtained 4 of the 10 enteric viruses detected. Picornaviruses, including hepatovirus, enterovirus, and cardiovirus, are an important group of viruses that display a diverse range of human infections and clinical symptoms. Hepatovirus was detected in all samples with the highest relative abundance among enteric viruses. Hepatitis A virus (HAV) is the type species of the hepatovirus genus and is the causative agent of hepatitis A (Lemon et al., 2018). Over 700 probable and confirmed cases of hepatitis A were reported in the service community during the 2017-2018 sampling years as a result of the 2016 multi-state hepatitis A outbreak (CDC, 2020a; MDHHS, 2020b). The prevalence of HAV in sequenced samples suggest the use of WBE for routine surveillance of hepatitis A outbreaks in communities.

Enteroviruses include enteric (poliovirus, coxsackievirus, echovirus) and respiratory (rhinovirus, enterovirus D68) pathogens (Wells and Coyne, 2019). Enteroviruses cause a host of illnesses including common cold, hand-foot-mouth disease, poliomyelitis, acute flaccid myelitis (AFM), and aseptic meningitis (Wells and Coyne, 2019). In 2018, there was a spike in AFM cases in the U.S. with Michigan obtaining 5 cases in that same year (CDC, 2020b). Although enteroviruses are not the only viruses that cause AFM, there have been well-established links (Dyda et al., 2018). Metagenomic analysis suggest the presence of polioviruses and other enterovirus species in wastewater samples during the 2017-2018 sampling period. This indicates the potential circulation of clinically important enteroviruses in the environment and potential connection to interruptions in community health.

Cardioviruses were believed to mainly infect rodents until 2007 when Saffold virus (SAFV), a novel cardiovirus, was identified in human stool. Since then SAFV has been identified in stool and

nasopharyngeal aspirates of patients suffering from gastrointestinal or respiratory illnesses. Since SAFV is commonly present in patients with coinfections, further investigations are needed to determine virus pathology. Moreover, SAFV has also been detected in Cerebrospinal fluid (CSF) of children, but these findings were not consistent across studies, even those focusing on patients with neurological disruptions (Tan et al., 2017). Nonetheless, SAFV associations with neuropathogenesis is of importance given its close relation to Theiler's Murine encephalomyelitis virus (TMEV), which causes neuropathogenesis in mice (Tan et al., 2017). Cardioviruses have been previously isolated from wastewater (Blinkova et al., 2009; Bonanno Ferraro et al., 2020) and were identified in 17 of 18 samples.

Apart from picornaviruses, mamastrovirus was the most prominent enteric virus genus detected in wastewater samples that primarily causes gastrointestinal illness (Bosch et al., 2014). Astroviruses are commonly detected in the environment during winter months (Bosch et al., 2014) and suggest the prevalence of astrovirus infections within the community during the sampling period. Although there were a significant number of cases linked to gastrointestinal illness during the sampling year (Table A3.3) it is difficult to assess the burden of astroviruses within the community given the presence of other viruses promoting similar clinical manifestations like norovirus and sapovirus.

Similar to enteric viruses, respiratory viruses were abundant in metagenomic samples with orthopoxivirus obtaining the greatest number of hits among the 26 human viral pathogens identified. Orthopoxiviruses contain respiratory pathogens like variola virus and pathogens transmitted through vaccination, zoonoses, or close contact such as vaccinia virus (Buller and Palumbo, 1991; Haller et al., 2014). Vaccinia virus (VACV) was the most prevalent species

detected within the orthopoxivirus genus. VACV has been used widely in human immunization against smallpox (Haller et al., 2014). Although routine vaccination against smallpox is no longer performed in the U.S., recommended vaccination is suggested for individuals who are at risk of exposure, for example, laboratory workers (CDC, 2017). The presence of VACV could be a result of viral shedding from recently vaccinated individuals, silent community spread, or environmental prevalence. Further investigation is needed to determine potential sources of VACV prevalence in the environment.

Moreover, betacoronavirus (BCoV) was detected 8 of 18 wastewater samples. BCoVs are known to cause respiratory illnesses in humans ranging from common cold to more severe diseases like Severe Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS), and Coronavirus Disease 2019 (COVID-19). Frequent mixing of human and animal reservoirs in densely populated areas facilitated outbreaks of viruses from this group in previous years including SARS-CoV in 2003 (Kuiken et al., 2003), MERS-CoV in 2012 (Chan et al., 2015), and SARS-CoV-2 in 2019 (Lai et al., 2020). Along with respiratory ailments, gastrointestinal illnesses have also been reported in patients with BCoV infections (Lai et al., 2020). Albeit viral and clinical data investigations for this study were conducted prior to the COVID-19 pandemic, two possible novel coronavirus cases were reported in the service community during the 2018 sampling year. Additionally, BCoV are known to cause flu-like symptoms and have therefore been linked to ILI reported in the service community. Like GI, the reporting of non-specific diseases poses limitations when determining the burden of such viral pathogens within communities. Routine testing for BCoVs in individuals displaying ILI infections is necessary to prevent large-scale outbreaks of known and novel coronaviruses.

Several clinically important bloodborne pathogens were also detected in sequenced samples. Bloodborne pathogens are often transmitted through contact with infected blood, bodily fluids, or indirect contact with contaminated fomites. Despite the introduction of vaccines and effective medical interventions, bloodborne pathogen hepatitis C virus (HCV) is the leading cause of liver disease worldwide (Chen and Morgan, 2006). HCV is the only species in the Hepacivirus genus known to infect humans. More than 9,000 probable cases of hepatitis C were reported in the service community during the 2017-2018 sampling year (Table A.3). Despite the high number of disease cases, HCV obtained a low detection rate (2/18). The primary route of transmission, location of viral shedding, duration of infection, and presence of environmental inhibitors can affect the presence of HCV in wastewater. Nonetheless, to our knowledge, this is the first study to detect hepacivirus in untreated wastewater from an urban community.

Along with hepacivirus, human associated lentivirus contigs were detected in 5 of 18 untreated wastewater samples. Lentiviruses are a group of retroviruses that cause chronic and often deadly diseases in vertebrates and are known to have long incubation periods. Human immunodeficiency virus (HIV) 1 and 2 are the only viruses contained in this group that cause infections in humans. HIV is the causative agent of Acquired immunodeficiency syndrome (AIDS). HIV has accounted for more than 30 million deaths since 1981 with the highest burden in southern Africa (del Rio, 2017). The virus is said to have originated from non-human primates through zoonosis and has since evolved to spread through human-to-human contact (Fox et al., 2002). HIV infections are more persistent within the study area as compared other counties (MDHHS, 2018b).

Other bloodborne pathogens include lymphocryptovirus, erythroparvovirus, and deltaretrovirus belonging to the Herpesviridae, Parvoviridae, and Retroviridae families, respectively. Lymphocryptovirus, containing the Epstein–Barrvirus (EBV), was the third most abundant genus in metagenomic samples. EBV is most known for causing mononucleosis (mono) commonly called the kissing disease. The virus is prevalent in > 90% of the world's population and is commonly transmitted through saliva but can also be spread via blood (Tsao et al., 2015). EBV infections are also known to cause epithelial cancers such as nasopharyngeal carcinoma (NPC) and EBV-associated gastric cancers (Tsao et al., 2015). Erythroparvovirus (B19V) is associated with fifth disease, which causes a rash primarily in children. It is transmitted mainly by the respiratory route often causing outbreaks in schools and day care centers. The virus can also be transmitted via blood and blood-contaminated fomites (Qiu et al., 2017). Viruses within the deltaretrovirus genus are recognized for causing Adult T-cell leukemia/lymphoma (ATL) and can be transmitted mother-to-child, sexual intercourse, and blood transfusions (Ishitsuka and Tamura, 2014). No mandatory reporting was required for primary infections associated with the abovementioned viruses during the study period.

Most samples contained contigs associated with vector-borne genus alphavirus. Alphaviruses consists of mainly viruses transmitted through insect bites (arthropod-borne). Nearly all arthropod-borne viruses are zoonotic (Weaver et al., 2018). Zoonotic viruses, viruses that spread from animals to humans, constitutes an important reservoir of viruses. Many clinically relevant vector-borne viruses are transmitted via mosquitoes including Zika, West Nile, Yellow Fever, Eastern equine encephalitis virus (EEEV), dengue, and chikungunya. Alphavirus consists of several vector-borne viruses including chikungunya virus (CHIKV), EEEV, and Venezuelan equine encephalitis virus (VEEV). Contigs assigned to human viral associated proteins in the alphavirus

genus were detected in 16 of 18 samples. EEEV and VEEV were the primary species in 3 and 12 samples, respectively.

EEEV and VEEV cause encephalitis and neurological impairment mainly in equine species and humans and were first identified in the United States in 1933 and in Venezuela in 1938, respectively (Armstrong and Andreadis, 2013; Weaver et al., 2004). Although EEEV infections in humans are rare compared to other clinically relevant vector-borne infections, it is the deadliest with a fatality rate of 35-75% (Armstrong and Andreadis, 2013). This contrasts with VEEV infections where fatalities rates of less than 1% have been reported (Weaver et al., 2004).

One case of EEE was reported in the service area during the 2018 sampling year. Similarly, the CDC reported a spiked in the number of EEE cases in 2019 with 38 confirmed cases and 15 deaths as of December 2019. Among the states affected were Michigan with 10 of the 38 cases (CDC, 2019b). Re-emerging infections of this nature are expected to rise in the future due to urbanization and climate change (McMichael, 2004). This study highlights the potential utility of wastewater to be used as a surveillance tool for vector-borne viruses. To our knowledge, no previous study has reported alphaviruses in raw wastewater samples. Potential exposure routes of alphaviruses into wastewater reservoirs could be facilitated through stormwater collection during rainfall events or viral shedding from infected individuals. Further studies are warranted to examine the fate of vector-borne viruses in wastewater systems.

4.2. Screening of Select Human Viral Pathogens

Quantitative analysis of select viruses was performed to strengthen metagenomic findings and assess viral concentrations in collected samples. HAV was detected in all samples during metagenomic analysis with significantly high viral loads in wastewater as compared to other viruses tested. The high occurrence of HAV concentration in wastewater could signify outbreak conditions and critical locations with increased spatial sampling. Along with HAV, HAdV, SaV, NoV GII, and HHV-6 were quantified in wastewater samples with positive detection rates in metagenomic samples. HHVs obtained the lowest concentrations among select viruses. HHV-6 and -8, also referred to as roseolovirus and Kaposi sarcoma (KS) associated herpesvirus, respectively, are non-enteric viruses which could explain their low concentrations in sewage samples as compared to the other viruses screened. HHV-8 was quantified in untreated wastewater using qPCR but went undetected in metagenomic analysis. HHV-8 is commonly transmitted through saliva and is said to be more pervasive among men who have sex with men (MSM) (Engels et al., 2007). Those with HIV are at greater risk for developing KS (Engels et al., 2007). Results here illustrate the utility of WBE to be used for enteric and non-enteric viruses. Furthermore, the integration of NGS with qPCR techniques provides a more sensitive application for low abundant viruses in wastewater systems.

5. Conclusion

- Metagenomics detected the presence of enteric and non-enteric viruses that cause clinically important diseases that were reported within the study area during the sampling year.
- Findings reveal evidence of re-emerging vector-borne viruses.
- Frequent and rigorous wastewater sampling along with integrative sample processing strategies can be employed to identify the etiological agent of non-specific diseases and viruses that poses a significant burden among inhabitants.

• Results presented in this study suggests that WBE has the potential to advance the area of disease outbreak mitigation and improve public health responses to large scale outbreaks and viral pandemics.

Acknowledgements

This study was funded by National Science Foundation Award #1752773. We thank Anil Gosine (Detroit Water and Sewerage Department) and Michael Jurban (Great Lakes Water Authority) for allowing access to the Detroit wastewater treatment utility and assisting with sampling. We thank the Research Technology Support Facility (Michigan State University) for assisting with sequencing and Professor Shinhan Shiu (Michigan State University) for assisting with bioinformatics analysis.

APPENDIX

APPENDIX A3: Supplementary Tables and Figures

Table A3.	1. Pri	mers	and	probes	for	select	viruses.

Virus	Abbreviation	Primer and Probe	Sequence $(5' \rightarrow 3')$	Final Conc.	Reference
Hepatitis A	HAV	Forward Reverse Probe	GGTAGGCTACGGGTGAAAC AACAACTCACCAATATCCGC FAM-CTTAGGCTAATACTTCTATGAAGAGATGC-BHQ-1	250 nM 250 nM 150 nM	Jothikumar et al., 2005
Sapovirus	SaV	SaV124F SaV1F SaV5F SaV1245R SaV124TP SaV5TP	GAYCASGCTCTCGCYACCTAC TTGGCCCTCGCCACCTAC TTTGAACAAGCTGTGGCATGCTAC CCCTCCATYTCAAACACTA FAM-CCRCCTATRAACCA-MGB-NQF FAM-TGCCACCAATGTACCA-MGB-NQF	400 nM 400 nM 400 nM 200 nM 200 nM	Oka et al., 2006
Norovirus GII	NoV GII	QNIF2d – Forward COG2R – Reverse QNIFs – Probe	ATGTTCAGRTGGATGAGRTTCTCWGA TCGACGCCATCTTCATTCACA FAM-AGCACGTGGGAGGGGCGATCG –TAMRA	500 nM 900 nM 250 nM	Le Guyader et al., 2009
Human Adenovirus (40, 41)	HAdV	HAdV-F4041-hex157f HAdV-F40-hex245r HAdV-F41-hex246r HAdV-F4041 hex214rprobe	ACCCACGATGTAACCACAGAC ACTTTGTAAGAGTAGGCGGTTTC CACTTTGTAAGAATAAGCGGTGTC FAM-CGACKGGCACGAAKCGCAGCGT-TAMRA	400 nM 200 nM 200 nM 300 nM	Xagoraraki et al., 2007
Human Herpesvirus-6	HHV-6	Taq1 forward Taq2 reverse H6S Probe	GACAATCACATGCCTGGATAATG TGTAAGCGTGTGGTAATGGACTAA FAM-AGCAGCTGGCGAAAAGTGCTGTGC-TAMRA	400 nM 400 nM 200 nM	Gautheret- Dejean et al., 2002
Human Herpesvirus-8	HHV-8	Forward Reverse Probe	CCGAGGACGAAATGGAAGTG GGTGATGTTCTGAGTACATAGCGG FAM-ACAAATTGCCAGTAGCCCACCAGGAGA-TAMRA	900 nM 900 nM 250 nM	Lallemand et al., 2000

Table A3. 2. Table A3.2. Summary of reads produced from sequenced cDNA samples and metagenomic alignment statistics. Viruses assigned to human viral group include Riboviruses and virus assigned to the root.

		Paw	Quality								
Sample Date	Interceptor	Sequences per paired-end	Sequences per paired-end	Contigs aligned	Virus- associated contigs	Proportion of Viral Contigs (%)	Viruses assigned to human viral groups	Putative human viral pathogens (Swiss-Prot)	Proportion of Putative Human Viral Pathogens of viral affiliated contigs (%)		
	O-NWI	41,295,408	39,432,740	207,093	168,203	81.22	7,269	329	0.20%		
17-Nov-17	NI-EA	32,998,405	31,982,651	200,084	164,008	81.97	6,590	317	0.19%		
	DRI	37,963,134	36,613,719	201,210	162,349	80.69	7,537	345	0.21%		
	O-NWI	36,157,930	34,391,548	155,588	124,584	80.07	4,934	103	0.08%		
1-Dec-17	NI-EA	35,910,020	34,741,068	128,731	98,660	76.64	4,468	176	0.18%		
	DRI	29,674,125	27,975,039	109,008	85,975	78.87	3,839	120	0.14%		
	O-NWI	34,865,428	33,777,286	200,223	163,026	81.42	5,991	82	0.05%		
14-Dec-17	NI-EA	45,032,334	43,728,453	262,134	213,185	81.33	8,034	116	0.05%		
	DRI	32,632,968	31,567,977	202,664	165,927	81.87	6,095	87	0.05%		
	O-NWI	41,474,809	39,201,580	201,521	162,671	80.72	6,456	156	0.10%		
19-Jan-18	NI-EA	45,009,092	43,068,418	245,561	198,568	80.86	7,569	178	0.09%		
	DRI	37,211,308	35,415,458	162,724	132,135	81.20	4,861	111	0.08%		
	O-NWI	22,518,125	21,492,470	118,521	97,294	82.09	3,665	75	0.08%		
2-Feb-18	NI-EA	20,854,987	19,372,755	116,059	94,466	81.39	3,665	93	0.10%		
	DRI	28,310,029	27,252,837	152,527	126,140	82.70	4,961	250	0.20%		
	O-NWI	40,197,247	36,913,009	73,772	52,947	71.77	3,574	155	0.29%		
16-Feb-18	NI-EA	37,750,571	35,515,787	67,867	49,703	73.24	2,823	199	0.40%		
	DRI	24,531,662	22,751,559	35,812	26,107	72.90	3,413	203	0.78%		
Total		624,387,582	595,194,354	2,841,099	2,285,948	1,431	95,744	3,095	3.27%		
Average		34,688,199	33,066,353	157,839	126,997	79	5,319	172	0.18%		
Range.Min		20,854,987	19,372,755	35,812	26,107	72	2,823	75	0.05%		
Range.Max		45,032,334	43,728,453	262,134	213,185	83	8,034	345	0.78%		
				2017					2018		
------------	----------------------------------	---------	-------	---------	--------	---------	---------	-------	---------	--------	---------
Group	Viral Disease	Detroit	Wayne	Oakland	Macomb	MI	Detroit	Wayne	Oakland	Macomb	MI
Foodborne	Norovirus	1	22	27	5	1075	0	14	141	6	1012
Influenza	Flu Like Disease* (ILI)	1309	27799	20542	28087	394,852	1057	27496	16097	20200	379,206
Influenza	Influenza	2240	4571	4301	4109	26018	4942	9168	8432	7562	47212
Influenza	Influenza, Novel	0	0	0	0	2	0	0	0	0	3
Meningitis	Meningitis - Aseptic	63	100	74	44	659	59	75	84	59	827
Other	Acute Flaccid Myelitis (AFM)	0	0	0	0	0	1	1	1	1	6
Other	Encephalitis, Post Chickenpox	0	0	0	0	0	1	0	0	0	1
Other	Encephalitis, Post Mumps	0	0	0	0	0	0	0	0	0	0
Other	Encephalitis, Post Other	0	5	2	1	21	1	1	0	3	10
Other	Gastrointestinal Illness	87	509	2453	6515	158,893	18	342	3339	12064	154,968
Other	Guillain-Barre Syndrome	2	9	6	8	53	2	6	6	11	47
Other	Hantavirus	0	1	0	0	0	0	0	0	0	0
Other	Hantavirus, Other	0	0	0	0	0	0	0	0	0	0
Other	Hantavirus, Pulmonary	0	0	0	0	0	0	0	0	0	0
Other	Hemorrhagic Fever	0	0	0	0	0	0	0	0	0	0
Other	Kawasaki	6	6	4	5	46	2	3	0	3	35
Other	Novel Coronavirus	0	0	0	0	0	0	2	0	0	1
Other	Reye Syndrome	0	0	0	0	0	0	0	0	0	0
Other	Encephalitis, Primary	1	1	1	4	15	2	0	0	2	18
Rabies	Rabies Animal	0	1	5	2	44	0	1	8	4	76
Rabies	Rabies Human	0	0	0	0	0	0	0	0	0	0

Table A3. 3. Reportable viral associated diseases in Michigan. Number of disease cases for Michigan, Detroit City, Wayne County, Oakland County, and Macomb County in 2017 and 2018. umber represent probable and confirmed cases.

Table A3. 3 (cont'd)

Vector- borne	Chikungunya~	0	1	0	1	4	0	0	1	0	2
Vector-		1	4	4	0	11	1	0	2	0	12
borne Voctor	Dengue Fever	-			-		-	-	_	-	
borne	California	0	0	0	0	0	0	0	1	0	2
Vector-	Encephalitis,	0	0	0	0	0	0	1	0	0	1
borne	Eastern Equine	0	0	0	0	0	0	1	0	0	1
Vector-	Encephalitis,	0	0	0	0	0	0	0	0	0	0
borne	Powassan Encenhalitis St										
borne	Louis	0	0	0	0	0	0	0	0	0	0
Vector-	Encephalitis,	0	0	0	0	0	0	0	0	0	0
borne	Western Equine	0	0	0	0	0	0	0	0	0	0
Vector-											
borne		6	3	5	8	43	9	26	12	11	105
(mosquito-		0	U	U	C			-0			100
borne)	West Nile Virus										
vector-	Vallow Favor	0	0	0	0	0	0	0	0	0	0
Vector-											
borne	Zika	1	3	0	1	10	1	1	0	0	2
Viral		120	104	90	100	705	20	01	20	25	222
Hepatitis	Hepatitis A	130	124	80	198	705	30	81	30	35	332
Viral		81	53	4	13	63	28	21	19	13	81
Hepatitis	Hepatitis B, Acute	01	55		15	05	20	21	17	15	01
Viral	Henry's D. Characia	292	325	202	116	1216	251	217	177	105	1080
Virol	Hepatitis B, Chronic										
V II al Henatitis	Perinatal	0	1	0	0	1	0	0	0	1	1
Viral	1 ermatar										. – .
Hepatitis	Hepatitis C, Acute	117	10	27	68	233	21	44	81	68	174
Viral	1	1021	1407	1022	011	11070	1477	1122	059	921	10240
Hepatitis	Hepatitis C, Chronic	1951	1427	1022	911	11970	14//	1152	938	851	10549
Viral	Hepatitis C,	0	0	1	1	2	0	0	1	1	11
Hepatitis	Perinatal	Ŭ	Ű		•	-	Ŭ	~ 		•	
Viral	Hepatitis C, Unknown*	0	0	0	0	31	NLR	NLR	NLR	NLR	NLR
riepaulis	UIKIIUWII										

Table A3. 3 (cont'd)

Viral Hepatitis	Hepatitis D	0	0	1	0	3	NLR	NLR	NLR	NLR	NLR
Viral Hepatitis	Hepatitis E	1	2	0	0	5	1	1	1	1	9
Viral Hepatitis	Hepatitis Non A Non B	0	0	0	0	0	NLR	NLR	NLR	NLR	NLR
Viral Hepatitis	Hepatitis - Unspecified	0	0	0	0	0	NLR	NLR	NLR	NLR	NLR
VPD	Chickenpox (Varicella)	29	56	84	31	543	11	23	60	40	436
VPD	Measles	0	0	0	1	2	0	0	10	0	19
VPD	Mumps	4	5	4	4	47	6	17	5	2	81
VPD	Polio	0	0	0	0	0	0	0	0	0	0
VPD	Rubella	0	1	0	0	0	0	0	0	0	0
VPD	Rubella - Congenital	0	0	0	0	0	0	0	0	0	0
VPD	Shingles	33	79	327	127	976	49	110	388	190	1252
VPD	**VZ Infection, Unspecified	10	183	1	4	270	3	71	1	8	212

Notes: Year-to-date (YTD) disease counts taken from 2017 and 2018 weekly disease reports (week 52), Annual MI taken from 2019 Weekly disease report (week 31)

Diseases no longer reported (NLR) are not included in the study, these diseases are highlighted in dark gray

"**Cases of infections with VZ virus that are unable to be classified as either Chickenpox or Shingles should be reported as VZ infection, unspecified" (MDHHS)



95% family-wise confidence level

Figure A3. 1. Tukey's post hoc statistical test results. Plot displays the comparison of mean virus concentrations. Intervals that lie on the dotted line mean that there is no difference between the mean concentrations between those two viruses. Intervals that lie to the left of the dotted line means that the concentration of the first virus listed in the pair is significantly less than the second virus. If interval is on the right side of the dotted line the first virus is significantly greater than the second. Horizontal lines represent the 95% confidence interval.

REFERENCES

REFERENCES

Andrews, S., 2010. FastQC. Babraham Bioinforma.

- Armstrong, P.M., Andreadis, T.G., 2013. Eastern equine encephalitis virus Old enemy, new threat. N. Engl. J. Med.
- Arvin, A.N.N.M., 1996. Varicella-Zoster Virus 9, 361–381.
- Barzon, L., Pacenti, M., Franchin, E., Pagni, S., Martello, T., Cattai, M., Cusinato, R., Palù, G., 2013. Excretion of west nile virus in urine during acute infection. J. Infect. Dis.
- Bibby, K., Viau, E., Peccia, J., 2011. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. Lett. Appl. Microbiol.
- Bisseux, M., Colombet, J., Mirand, A., Roque-Afonso, A.M., Abravanel, F., Izopet, J., Archimbaud, C., Peigue-Lafeuille, H., Debroas, D., Bailly, J.L., Henquell, C., 2018. Monitoring human enteric viruses in wastewater and relevance to infections encountered in the clinical setting: A one-year experiment in central France, 2014 to 2015. Eurosurveillance 23, 1–11.
- Blinkova, O., Rosario, K., Li, L., Kapoor, A., Slikas, B., Bernardin, F., Breitbart, M., Delwart, E., 2009. Frequent detection of highly diverse variants of Cardiovirus, Cosavirus, Bocavirus, and Circovirus in sewage samples collected in the United States. J. Clin. Microbiol. 47, 3507– 3513.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., Schneider, M., 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 31, 365–370.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics.
- Bonanno Ferraro, G., Mancini, P., Veneri, C., Iaconelli, M., Suffredini, E., Brandtner, D., La Rosa, G., 2020. Evidence of Saffold virus circulation in Italy provided through environmental surveillance. Lett. Appl. Microbiol. 70, 102–108.
- Bosch, A., Pintó, R.M., Guix, S., 2014. Human astroviruses. Clin. Microbiol. Rev. 27, 1048–1074.
- Breitwieser, F.P., Lu, J., Salzberg, S.L., 2018. A review of methods and databases for metagenomic classification and assembly. Brief. Bioinform. 20, 1125–1139.

Buller, R.M.L., Palumbo, G.J., 1991. Poxvirus pathogenesis. Microbiol. Rev. 55, 80-122.

- Burns, M., Valdivia, H., 2008. Modelling the limit of detection in real-time quantitative PCR. Eur. Food Res. Technol. 226, 1513–1524.
- Cantalupo, P.G., Calgua, B., Zhao, G., Hundesa, A., Wier, A.D., Katz, J.P., Grabe, M., Hendrix, R.W., Girones, R., Wang, D., Pipas, J.M., 2011. Raw sewage harbors diverse viral populations. MBio 2, 1–11.
- CDC, 2017. Who should get vaccination [WWW Document]. URL https://www.cdc.gov/smallpox/vaccine-basics/who-gets-vaccination.html#care-for (accessed 4.27.20).
- CDC, 2019a. U.S. Influenza Surveillance System: Purpose and Methods [WWW Document]. FluView. URL https://www.cdc.gov/flu/weekly/overview.htm (accessed 4.28.20).
- CDC, 2019b. Eastern equine encephalitis [WWW Document]. URL https://www.cdc.gov/easternequineencephalitis/index.html (accessed 4.27.20).
- CDC, 2020a. Widespread person-to-person outbreaks of hepatitis A across the United States [WWW Document]. Viral Hepat. URL https://www.cdc.gov/hepatitis/outbreaks/2017March-HepatitisA.htm (accessed 4.27.20).
- CDC, 2020b. AFM cases and outbreaks | CDC [WWW Document]. URL https://www.cdc.gov/acute-flaccid-myelitis/cases-in-us.html (accessed 4.27.20).
- Chan, J.F.W., Lau, S.K.P., To, K.K.W., Cheng, V.C.C., Woo, P.C.Y., Yue, K.Y., 2015. Middle East Respiratory syndrome coronavirus: Another zoonotic betacoronavirus causing SARSlike disease. Clin. Microbiol. Rev. 28, 465–522.
- Chen, S.L., Morgan, T.R., 2006. The natural history of hepatitis C virus (HCV) infection. Int. J. Med. Sci. 3, 47–52.
- De Bolle, L., Naesens, L., De Clercq, E., 2005. Update on human herpesvirus 6 biology, clinical features, and therapy. Clin. Microbiol. Rev. 18, 217–245.
- de Crom, S.C.M., Rossen, J.W.A., van Furth, A.M., Obihara, C.C., 2016. Enterovirus and parechovirus infection in children: a brief overview. Eur. J. Pediatr. 175, 1023–1029.
- del Rio, C., 2017. The global HIV epidemic: What the pathologist needs to know. Semin. Diagn. Pathol. 34, 314–317.
- Denner, J., 2010. Detection of a gamma etrovirus, XMRV, in the human population: Open questions and implications for xenotransplantation. Retrovirology.
- Doorbar, J., Egawa, N., Griffin, H., Kranjec, C., Murakami, I., 2015. Human papillomavirus molecular biology and disease association. Rev. Med. Virol.

- Dyda, A., Stelzer-Braid, S., Adam, D., Chughtai, A.A., Macintyre, C.R., 2018. The association between acute flaccid myelitis (AFM) and enterovirus D68 (EV-D68) what is the evidence for causation? Eurosurveillance 23, 1–9.
- Engels, E.A., Atkinson, J.O., Graubard, B.I., McQuillan, G.M., Gamache, C., Mbisa, G., Cohn, S., Whitby, D., Goedert, J.J., 2007. Risk factors for human herpesvirus 8 infection among adults in the United States and evidence for sexual transmission. J. Infect. Dis. 196, 199–207.
- Fancello, L., Trape, S., Robert, C., Boyer, M., Popgeorgiev, N., Raoult, D., Desnues, C., 2013. Viruses in the desert: A metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. ISME J. 7, 359–369.
- Farkas, K., Cooper, D.M., McDonald, J.E., Malham, S.K., de Rougemont, A., Jones, D.L., 2018. Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine receiving waters. Sci. Total Environ. 634, 1174–1183.
- Fox, J.G., Newcomer, C.E., Rozmiarek, H., 2002. Selected zoonoses, Second Edi. ed, Laboratory Animal Medicine. Elsevier Inc.
- Gautheret-Dejean, A., Manichanh, C., Thien-Ah-Koon, F., Fillet, A.M., Mangeney, N., Vidaud, M., Dhedin, N., Vernant, J.P., Agut, H., 2002. Development of a real-time polymerase chain reaction assay for the diagnosis of human herpesvirus-6 infection and application to bone marrow transplant patients. J. Virol. Methods 100, 27–35.
- Ghebremedhin, B., 2014. Human adenovirus: Viral pathogen with increasing importance. Eur. J. Microbiol. Immunol. 4, 26–33.
- Glass, R.I., Parashar, U.D., Estes, M.K., 2009. Norovirus gastroenteritis. N. Engl. J. Med.
- GLWA, 2018. Capital Improvement Plan 2019-2023.
- Gonçalves, D.U., Proietti, F.A., Ribas, J.G.R., Araújo, M.G., Pinheiro, S.R., Guedes, A.C., Carneiro-Proietti, A.B.F., 2010. Epidemiology, treatment, and prevention of human T-cell leukemia virus type 1-associated diseases. Clin. Microbiol. Rev.
- Gourinat, A.C., O'Connor, O., Calvez, E., Goarant, C., Dupont-Rouzeyrol, M., 2015. Detection of zika virus in urine. Emerg. Infect. Dis.
- Gundy, P.M., Gerba, C.P., Pepper, I.L., 2009. Survival of coronaviruses in water and wastewater. Food Environ. Virol.
- Hall, C.B., Long, C.E., Schnabel, K.C., Caserta, M.T., Mcintyre, K.M., Costanzo, M.A., Knott, A., Dewhurst, S., Insel, R.A., Epstein, L.G., 1994. Human herpesvirus-6 infection in children A prospective study of complications and reactivation. N. Engl. J. Med.

Haller, S.L., Peng, C., McFadden, G., Rothenburg, S., 2014. Poxviruses and the evolution of host

range and virulence. Infect. Genet. Evol.

- Heijnen, L., Medema, G., 2011. Surveillance of influenza A and the pandemic influenza A (H1N1) 2009 in sewage and surface water in the Netherlands. J. Water Health 9, 434–442.
- Hirayama, T., Mizuno, Y., Takeshita, N., Kotaki, A., Tajima, S., Omatsu, T., Sano, K., Kurane, I., Takasaki, T., 2012. Detection of dengue virus genome in urine by real-time reverse transcriptase PCR: A laboratory diagnostic method useful after disappearance of the genome in serum. J. Clin. Microbiol.
- Hirose, R., Daidoji, T., Naito, Y., Watanabe, Y., Arai, Y., Oda, T., Konishi, H., Yamawaki, M., Itoh, Y., Nakaya, T., 2016. Long-term detection of seasonal influenza RNA in faeces and intestine. Clin. Microbiol. Infect. 22, 813.e1-813.e7.
- Ishitsuka, K., Tamura, K., 2014. Human T-cell leukaemia virus type I and adult T-cell leukaemialymphoma. Lancet Oncol. 15, e517–e526.
- Jones, B., Cushingberry, G Jr., Ayers J., Benson, S., Castaneda-Lopez, R., Leland, G., Sheffield, M., A.S., 2015. Rehabilitation of the rectangular primary clarifiers, electrical/mechanical buildings and pipe gallery. Detroit.
- Jothikumar, N., Cromeans, T.L., Sobsey, M.D., Robertson, B.H., 2005. Development and evaluation of a broadly reactive TaqMan assay for rapid detection of hepatitis A virus. Appl. Environ. Microbiol. 71, 3359–3363.
- Kamar, N., Dalton, H.R., Abravanel, F., Izopet, J., 2014. Hepatitis E virus infection. Clin. Microbiol. Rev. 27, 116–138.
- Kuiken, T., Fouchier, R.A.M., Schutten, M., Rimmelzwaan, G.F., Van Amerongen, G., Van Riel, D., Laman, J.D., De Jong, T., Van Doornum, G., Lim, W., Ling, A.E., Chan, P.K.S., Tam, J.S., Zambon, M.C., Gopal, R., Drosten, C., Van Der Werf, S., Escriou, N., Manuguerra, J.C., Stöhr, K., Peiris, J.S.M., Osterhaus, A.D.M.E., 2003. Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. Lancet 362, 263–270.
- Lai, C.C., Shih, T.P., Ko, W.C., Tang, H.J., Hsueh, P.R., 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int. J. Antimicrob. Agents 55, 105924.
- Lallemand, F., Desire, N., Rozenbaum, W., Nicolas, J.C., Marechal, V., 2000. Quantitative analysis of human herpesvirus 8 viral load using a real- time PCR assay. J. Clin. Microbiol. 38, 1404–1408.
- Le Guyader, F.S., Parnaudeau, S., Schaeffer, J., Bosch, A., Loisy, F., Pommepuy, M., Atmar, R.L., 2009. Detection and quantification of noroviruses in shellfish. Appl. Environ. Microbiol. 75, 618–624.

- Lemon, S.M., Ott, J.J., Van Damme, P., Shouval, D., 2018. Type A viral hepatitis: A summary and update on the molecular virology, epidemiology, pathogenesis and prevention. J. Hepatol. 68, 167–184.
- Li, T., Mbala-Kingebeni, P., Naccache, S.N., Thézé, J., Bouquet, J., Federman, S., Somasekar, S., Yu, G., Martin, C.S.S., Achari, A., Schneider, B.S., Rimoin, A.W., Rambaut, A., Nsio, J., Mulembakani, P., Ahuka-Mundeke, S., Kapetshi, J., Pybus, O.G., Muyembe-Tamfum, J.J., Chiu, C.Y., 2019. Metagenomic next-generation sequencing of the 2014 ebola virus disease outbreak in the democratic republic of the Congo. J. Clin. Microbiol.
- Mawson, A.R., Croft, A.M., 2019. Rubella virus infection, the congenital rubella syndrome, and the link to autism. Int. J. Environ. Res. Public Health 16.
- McCall, C., Xagoraraki, I., 2019. Metagenomic approaches for detecting viral diversity in water environments. J. Environ. Eng. 145.
- McMichael, A.J., 2004. Environmental and social influences on emerging infectious diseases: Past, present and future. In: Philosophical Transactions of the Royal Society B: Biological Sciences.
- MDHHS, 2018a. Managing communicable diseases in schools [WWW Document]. URL https://www.michigan.gov/documents/mdch/Managing_CD_in_Schools_FINAL_469824_7 .PDF (accessed 4.28.20).
- MDHHS, 2018b. Epidemiologic profile of HIV in Michigan.
- MDHHS, 2020a. Michigan Disease Surveillance System background [WWW Document]. URL https://www.michigan.gov/mdhhs/0,5885,7-339-71550_5104_31274-96814--,00.html (accessed 4.28.20).
- MDHHS, 2020b. Michigan hepatitis A outbreak: Update [WWW Document]. URL https://www.michigan.gov/mdhhs/0,5885,7-339-71550_2955_2976_82305_82310-447907--,00.html (accessed 4.27.20).
- Miranda, J.A., Culley, A.I., Schvarcz, C.R., Steward, G.F., 2016. RNA viruses as major contributors to Antarctic virioplankton. Environ. Microbiol. 18, 3714–3727.
- Munir, M., Wong, K., Xagoraraki, I., 2011. Release of antibiotic resistant bacteria and genes in the effluent and biosolids of five wastewater utilities in Michigan. Water Res.
- Ng, T.F.F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B.S., Wommack, K.E., Delwart, E., 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J. Virol. 86, 12161–12175.
- O'Brien, E., Nakyazze, J., Wu, H., Kiwanuka, N., Cunningham, W., Kaneene, J.B., Xagoraraki, I., 2017. Viral diversity and abundance in polluted waters in Kampala, Uganda. Water Res.

- O'Brien, E., Xagoraraki, I., 2019. A water-focused one-health approach for early detection and prevention of viral outbreaks. One Heal.
- Oka, T., Katayama, K., Hansman, G.S., Kageyama, T., Ogawa, S., Wu, F.T., White, P.A., Takeda, N., 2006. Detection of human sapovirus by real-time reverse transcription-polymerase chain reaction. J. Med. Virol.
- Oka, T., Wang, Q., Katayama, K., Saifb, L.J., 2015. Comprehensive review of human sapoviruses. Clin. Microbiol. Rev. 28, 32–53.
- Poloni, T.R., Oliveira, A.S., Alfonso, H.L., Galvo, L.R., Amarilla, A.A., Poloni, D.F., Figueiredo, L.T., Aquino, V.H., 2010. Detection of dengue virus in saliva and urine by real time RT-PCR. Virol. J.
- Qiu, J., Söderlund-Venermo, M., Young, N.S., 2017. Human parvoviruses. Clin. Microbiol. Rev.
- R Core Team, 2019. R: A language and environment for statistical computing. Vienna, Austria.
- Tan, S.Z.K., Tan, M.Z.Y., Prabakaran, M., 2017. Saffold virus, an emerging human cardiovirus. Rev. Med. Virol. 27.
- Tonry, J.H., Brown, C.B., Cropp, C.B., Co, J.K.G., Bennett, S.N., Nerurkar, V.R., Kuberski, T., Gubler, D.J., 2005. West Nile virus detection in urine. Emerg. Infect. Dis.
- Tsao, S.W., Tsang, C.M., To, K.F., Lo, K.W., 2015. The role of Epstein-Barr virus in epithelial malignancies. J. Pathol. 235, 323–333.
- U.S. EPA, 2001. Manual of Methods for Virology (Chapter 14).
- Van Den Berg, B., Walgaard, C., Drenthen, J., Fokke, C., Jacobs, B.C., Van Doorn, P.A., 2014. Guillain-Barré syndrome: Pathogenesis, diagnosis, treatment and prognosis. Nat. Rev. Neurol. 10, 469–482.
- Victoria, M., Tort, L.F.L., García, M., Lizasoain, A., Maya, L., Leite, J.P.G., Miagostovich, M.P., Cristina, J., Colina, R., 2014. Assessment of gastroenteric viruses from wastewater directly discharged into Uruguay River, Uruguay. Food Environ. Virol. 6, 116–124.
- Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., Mardis, E.R., Hickenbotham, M., Magrini, V., Eldred, J., Latreille, J.P., Wilson, R.K., Ganem, D., DeRisi, J.L., 2003. Viral discovery and sequence recovery using DNA microarrays. PLoS Biol.
- Weaver, S.C., Charlier, C., Vasilakis, N., Lecuit, M., 2018. Zika, chikungunya, and other emerging vector-borne viral diseases. Annu. Rev. Med.
- Weaver, S.C., Ferro, C., Barrera, R., Boshell, J., Navarro, J.-C., 2004. V Enezuelan E Quine E Ncephalitis . Annu. Rev. Entomol. 49, 141–174.

- Wells, A.I., Coyne, C.B., 2019. Enteroviruses: A gut-wrenching game of entry, detection, and evasion. Viruses 11, 1–20.
- Whitley, R.J., 2011. Herpes Simplex Virus Infections. Goldman's Cecil Med. Twenty Fourth Ed. 2, 2125–2128.
- Widener, R.W., Whitley, R.J., 2014. Herpes simplex virus, 1st ed, Handbook of Clinical Neurology. Elsevier B.V.
- Xagoraraki, I., Kuo, D.H.W., Wong, K., Wong, M., Rose, J.B., 2007. Occurrence of human adenoviruses at two recreational beaches of the great lakes. Appl. Environ. Microbiol.
- Xagoraraki, I., O'Brien, E., 2020. Wastewater-based epidemiology for early detection of viral outbreaks.

CHAPTER 4: ASSESSMENT OF CALICIVIRUSES AND OTHER ENTERIC VIRUSES IN A LARGE METROPOLITAN AREA USING WASTEWATER SURVEILLANCE

Submitted, in part, for publication: Camille McCall, Huiyun Wu, Irene Xagoraraki

Abstract

It is well known that caliciviruses are leading causes of acute gastroenteritis (AGE) globally. Although the presence of noroviruses (NoVs) in wastewater and clinical settings is well studied, further work to understand the presence of sapoviruses (SaVs) in wastewater and burden on AGE is warranted, particularly in the United States. The present study investigates NoV GII and SaV in wastewater in comparison to clinical gastrointestinal cases in a large metropolitan area in the United States over the course of two winters. Metagenomics analysis was performed to characterized NoV and SaV genotypes in collected samples and screen for other enteric viruses, specifically those causing diarrheal illnesses. Average NoV GII and SaV concentrations in wastewater for the first sampling year were 1.36×10^6 gc/L and 2.94×10^4 gc/L, respectively. The second year of sampling resulted in average concentrations of 1.34×10^{6} gc/L for SaV and 3.55×10^{4} gc/L for NoV GII. There was no significant correlation between calicivirus concentrations in wastewater and the number of gastrointestinal and noroviruses cases reported in the catchment area. NoV genogroup GI and GII were detected in wastewater and SaV GI.1 was the only genotype detected using metagenomics. Adenovirus, astrovirus, parechovirus, norovirus, sapovirus and bocavirus were among the 9 enteric viruses detected by metagenomics and are commonly associated with diarrhea-related illnesses. Findings presented suggest a greater burden of SaV infections in the catchment area as compared to NoV infections. The presence of other pathogens causing diarrheal illnesses may contributed more to gastrointestinal cases during the time of sampling. Routine monitoring and reporting of sapovirus infections and other enteric viruses, like astrovirus can improve wastewater surveillance approaches. Furthermore, findings from this study demonstrate the usefulness of metagenomics for genogrouping and viral surveillance.

1. Introduction

Noroviruses (NoVs) are among the leading causes of outbreaks of viral acute gastroenteritis (AGE) worldwide (Glass et al., 2009; Hall et al., 2013). Although not as notorious as NoV, sapoviruses (SaVs) are also important pathogens in AGE cases. NoV and SaV are nonenveloped positive-sense single-stranded RNA viruses in the Caliciviridae family. Both viruses are considered enteric pathogens that infect the intestinal tract and can be spread via the fecal-to-oral route or person-toperson. NoVs contain at least seven different genogroups (GI, GII, GII, GIV, GV, GVI, GVII) where GI and GII are most commonly isolated from human clinical specimens. NoV GII predominates in sporadic cases and outbreaks globally (Glass et al., 2009). Genetic sequences of NoV samples from outbreak investigation and sporadic cases occurring between 2005 and 2016 identified 92% of sequences as belonging to NoV GII, less than 10% of sequences were classified as GI and less than 1% ad GIV (van Beek et al., 2018). Similarity, Cannon et al., 2017 found NoV GII to be responsible for approximately 82% of norovirus outbreak in the United Stated between 2013 and 2016. SaVs include at least five confirmed groups where GI, GII, GIV, and GV are known to infect humans (Oka et al., 2015). GI has been described in several outbreak cases (Kumthip et al., 2018; Sánchez et al., 2018).

Monitoring of NoV in wastewater has provided insight into the transmission of these clinically important viruses and detected correlation between NoV-related AGE outbreaks and virus circulation in the environment (Farkas et al., 2018a; Hellmér et al., 2014; Iwai et al., 2009). Although our understanding of sapovirus prevalence in communities is expanding, little work has been done to understand the burden and environmental distribution of sapoviruses in the United States (Kitajima et al., 2018). Additionally, the presence of other enteric viruses can be screened in wastewater using metagenomics to provide insight into causes of unspecified AGE cases and prevent widespread outbreaks.

The authors quantified SaV and NoV GII in wastewater samples collected from a large metropolitan area in the United States over the course of two winters. A correlation analysis was done to evaluate the relationship between gastrointestinal illnesses and norovirus cases reported within the catchment area and viral concentrations. NoV and SaV genogroups were characterized and samples were screened for other enteric viruses commonly causing AGE using metagenomics. This study builds on current knowledge of calicivirus prevalence in the United States and addresses the need for routine clinical monitoring of SaV in communities.

2. Methods

2.1. Study Area and Wastewater Sample Collection

Wastewater samples were collected from the Water Resource Recovery Facility (WRRF) located in Detroit, Michigan. The Detroit WRRF is the largest single site wastewater treatment plant in the U.S. and treats wastewater from an estimated 3 million inhabitants with an average daily flow of 650 MGD (GLWA, 2018). It services the three largest counties, by population, in Michigan. These are Wayne, Oakland, and Macomb counties (Jones et al., 2015). The WRRF receives wastewater from its service municipalities via three main interceptors (sewers): North Interceptor-East Arm (NI-EA), Detroit River Interceptor (DRI), and Oakwood-Northwest-Wayne County Interceptor (O-NWI).

Untreated wastewater samples were collected at the WRRF from sampling points located at each of the three interceptors approximately bi-weekly between November 2017 and February 2018 (n=54) for sampling year one (SY1) and October 2018 through March 2019 for sampling year two (SY2) (n=58). Due to operational conditions during SY2, DRI and O-NWI sampling sites were not sampled on all occasions. Viruses were isolated from untreated wastewater using electropositive NanoCeram column filters following the EPA's virus adsorption-elution protocol (U.S. EPA, 2001). Sewage samples were collected in triplicates for each interceptor with average filtered volumes ranging between 24-44 liters per interceptor. Each interceptor was sampled with its own filter house, tubing, and vacuum pump to minimize cross contamination. Virus filters were immediately stored on ice and transported to the Environmental Virology Laboratory at Michigan State University (MSU) and stored in -20°C until further processing.

2.2. Sample Processing and Virus Isolation

Following wastewater sampling, NanoCeram cartridge filters were eluted within 24 h with 1.5% w/v beef extract (0.05 M glycine, pH 9.5) according to the EPA's protocol (U.S. EPA, 2001). In short, filters were eluted with 1 L of beef extract for a total of 2 min. The pH of the solution was adjusted to 3.5 ± 0.1 and flocculated for 30 min before centrifugation at 2500g for 15 min at 4°C. Supernatant was discarded and pellets were resuspended in 30 mL of 0.15 M sodium phosphate (pH 9.0-9.5) followed by a second round of centrifugation carried out at 7000g for 10 min at 4°C.

The supernatant was neutralized (pH ~7.25) and subjected to filtration using to 0.45 μ m and 0.22 μ m syringe filters to eliminate bacterial contamination. Extraction of nucleic acid was performed on 140 μ L of purified virus concentrate using the QIAamp Viral RNA Mini Kit (Qiagen) following the manufacturer's protocol and eluted in 80 μ L of elution buffer. Nucleic acid was stored at -80°C until further processing.

2.3. Metagenomic Analysis

2.3.1. Random Amplification and Next Generation Sequence Processing

Purified nucleic acid from each biological replicate was pooled together for a total of 18 samples representing genetic material from all three interceptors during each of the six sampling dates. Nucleic acid from each sample was reverse transcribed and subjected to random amplification as previously described (Wang et al., 2003) to evaluate both RNA and DNA viruses. Eighteen samples of viral cDNA were sent to the Research Technology Support Facility Genomics Core at Michigan State University for whole-genome shotgun sequencing (WGS). The Illumina TruSeq Nano DNA Library Preparation Kit was used for all cDNA samples. Library preparation was performed on a Perkin Elmer Sciclone G3 robot according to the manufacturer's recommendations. This was followed by sequencing on an Illumina HiSeq4000 platform generating 150 bp paired-end reads.

2.3.2. Sequence Analysis and Taxonomic Anotation

Sequencing reads generated from WGS were processed on a Unix system through the MSU High Performance Computing Center (HPCC). Raw sequences were analyzed for quality using FastQC, a quality control tool for sequencing data (Andrews, 2010). Sequencing adapters and reads with an average quality score below 20 were removed using Trimmomatic (Bolger et al., 2014). Trimmed reads were assembled with IDBA-UD, a short-read de novo sequence aligner for metagenomic data. Reads were assembled into contigs using an iterative k-mer approach with kmer sizes ranging between 40 and 120 in increments of 10. The remaining parameters were run at default conditions.

Since viral genomes can be difficult to detect in metagenomic datasets, an optimized multialignment approach was used to improve alignment and annotation of viral reads. First, contigs were aligned against the Viral RefSeq database using tBLASTx with an E-value of 10⁻³. This approach has been known to increase human viral discovery in metagenomic datasets (Bibby et al., 2011). Aligned contigs were assigned to the lowest common ancestor (LCA) according to the NCBI's taxonomy with MEGAN (v. 6.15.0). The top 10 percent of BLAST alignments with a minimum bit score of 50 and contig coverage of at least 80% were considered in taxonomic analysis. The remaining parameters were run at default conditions. Contigs assigned to human virus groups were extracted for further analysis of enteric viruses. Extracted contigs were aligned with BLASTx with an E-value of 10⁻⁵ against a custom human virus database containing 5,979 human viral proteins in Swiss-Prot database (Boeckmann et al., 2003). These sequences represented all human viruses in the Swiss-Prot database at the time of retrieval (September 2019). Contigs annotated as NoV and SaV were genotyped using the Norovirus Typing Tool 2.0 (https://www.rivm.nl/mpf/typingtool/norovirus/). This tool is a reliable web-based genotyping platform, which performs a series of steps including BLAST alignment along with phylogenic analysis and bootstrap validation to characterize caliciviruses and enterovirus genotypes (Kroneman et al., 2011).

2.4. Quantitative PCR

Quantitative synthetic NoV GII RNA and SaV RNA was obtained from ATCC. RNA was diluted 10-fold and analyzed as described in the following section. Standard curves for NoV GII and SaV obtained R-squared values of > 99% and slopes of -3.82 and -3.35, respectively. The limit of detection (LOD) for each virus was determined by the lowest point on the standard curve. NoV GII and SaV obtained detection rates of 10^1 gc/µL. Quantitative PCR was used to establish the concentration of NoV GII and SaV in wastewater samples. All qPCRs were performed in triplicates on a Mastercycler ep realplex2 (Eppendorf) in 96-well optical plates. Amplification of cDNA was mediated using Lightcycler 480 Probes Master (Roche) at a concentration of $1 \times$ in all reactions. Sterile nuclease free water was used to meet volume requirements in all reactions.

SaV was quantified using a two-step RT-qPCR based on a previously described method (Oka et al., 2006). Briefly, viral RNA was reverse transcribed using iScript RT-qPCR Supermix (Bio-Rad) according to the manufacturer's protocol. SaV quantification was carried out in a 25 uL reaction containing each primer and probe. Reactions were performed with the following conditions: 95°C for 15 min, followed by 45 cycles of 94°C for 15 s, 62°C for 1 min, and 72°C for 15 s. Norovirus GII was quantified using a one-step RT-qPCR as previously described (Le Guyader et al., 2009). In short, the RT-qPCR was carried out in a 25 µL reaction mixture containing primers and probe,

2 μL of iScript RT-qPCR Supermix, and 5 μL of viral RNA, negative control, or positive control. Reactions were performed with the following conditions: reverse transcription at 25°C for 5 min, 46°C for 20 min, and 95°C for 1 min, followed by 45 cycles of 95°C for 15 s, 60°C for 1 min, 65°C for 1 min.

2.5. Health Data Collection

The number of cases for norovirus and gastrointestinal illness (GI) in each service county was obtained from the Michigan Department of Health and Human Services (MDHHS). Probable and confirmed case counts were extracted from the Michigan Disease and Surveillance System (MDSS) weekly surveillance reports (WSR). The MDSS is a communicable disease reporting system used to facilitate coordination and sharing of disease surveillance data among multiple shareholders including healthcare providers and medical laboratories (MDHHS, 2020). Each weekly surveillance report accounts for disease cases reported from Sunday-Saturday of the corresponding week. It is important to note that the WSR uses GI to represent any disease displaying symptoms of this nature. The etiological agent of the disease is unspecified, but could be of viral, bacterial, or parasitic origin. GI is defined as symptoms related to diarrhea and/or vomiting (MDHHS, 2018).

2.6. Statistical Analysis

A one-way analysis of variance (ANOVA) and Tukey's HSD post hoc tests were used to investigate significance between mean concentrations of viruses in wastewater samples. Finally, Spearman's correlation analysis was used to determine significant associations between caliciviruses and the number of norovirus and GI reported cases within the service community. All statistical analyses were performed in R (R Core Team, 2019).

3. Results

3.1. Calicivirus detection in Wastewater

A temporal investigation of NoV GII and SaV concentrations in wastewater samples was carried out to compare the burden of these viruses and their contribution to related diseases within the service community. SaV detection rates in wastewater were 94.4% for SY1 and 85% for SY2. Likewise, NoV detection rates were 100% and 89% for SY 1 and SY 2, respectively. Concentrations of SaV for SY1 range between 1.10x10⁵ and 4.66x10⁶ gc/L (1.36x10⁶ gc/L) and 8.78Ex10⁴ and 5.20x10⁶ gc/L (1.34x10⁶ gc/L) for SY2. SaV concentrations were significantly greater than NoV GII concentrations, which were $1.16\text{Ex}10^3$ to $1.15\text{x}10^5$ gc/L ($2.94\text{x}10^4$ gc/L) and 7.22×10^2 to 2.01×10^5 gc/L (3.55×10^4 gc/L) for SY1 and SY2, respectively (P < 0.0001) (Figure 4.1). There was no significant difference in calicivirus concentrations between interceptors or sampling year (P > 0.05). Although SaV concentration remained largely the same between dates within the same year, NoV concentrations varied significantly (P < 0.0001). Generally, concentrations of NoV GII varied considerably in SY2 with increased concentrations observed during the latter part of the sampling year as illustrated in Figure 4.1. According to Spearman's correlation analysis, there was no significant association between the number of reported GI cases and concentrations of NoV GII and SaV per sampling week for either year (Table 4.1).



Figure 4. 1. Boxplots of norovirus genogroup II (NoV GII) and sapovirus (SaV) average concentrations in wastewater samples per sampling date during years one (A) and two (B). Median concentrations are denoted with a horizontal line.

Table 4. 1. Average concentrations of NoV GII and SaV per sampling date along with number of GI and noroviruses cases reported within the service community.

Sampling Date	Week's Total No. of Norovirus and GI Cases	Average SaV concentration (copies/L)	Spearman (SaV)	Average NoV GII concentration (copies/L)	Spearman (NoV GII)
17-Nov-17	0	6.57E+05	P = 0.42	1.06E+04	P = 0.54
1-Dec-17	5	2.18E+06	rho (ρ) = -0.41	2.03E+04	rho (ρ) = -0.32
14-Dec-17	0	2.38E+06		6.86E+04	
19-Jan-18	23	2.40E+05		3.14E+04	
2-Feb-18	25	2.28E+06		4.88E+04	
16-Feb-18	48	4.66E+05		3.93E+03	
17-Oct-18	882	Х	P = 0.79	Х	P = 0.05
31-Oct-18	0	1.91E+06	rho (ρ) = -0.13	7.22E+02	rho (ρ) = 0.71
28-Nov-18	0	3.41E+05		2.12E+03	
12-Dec-18	1091	7.28E+05		5.10E+03	
17-Jan-19	933	2.63E+06		7.64E+04	
7-Feb-19	0	х		2.55E+04	
14-Feb-19	1407	3.99E+05		3.30E+04	
28-Feb-19	15	1.56E+06		1.50E+04	
14-Mar-19	1621	1.30E+06		1.38E+05	

3.2. Metagenomic Screening of NoV and SaV

Illumina sequencing was performed on 18 untreated wastewater samples collected from Detroit's WRRF from November 2017 to February 2018. A total of 624.4 million reads were subject to quality trimming resulting in 595.2 million reads. The proportion of contigs assigned to viral taxonomic groups range between 72-83% with an average of approximately 0.18% (0.05-0.78%) affiliated to human viral taxa. NoV and SaV related contigs were detected in 50% and 28% of metagenomic samples, respectively (Figure 4.2). Protein sequences related to NoV GII (NCBI accession no. P54634) and SaV GI (NCBI accession no. Q69014) were identified with > 90% identity in 8/9 and 4/5 samples, respectively. NoV GI (accession no. Q04544) and GII were both detected in the NI-EA interceptor during the 14-Dec-17 sampling date. SaV GI contigs were typed as GI.1.



Figure 4. 2. Normalized abundance of norovirus and sapovirus in metagenomic samples.

Metagenomic analysis detected several other enteric viruses in wastewater samples belonging to the *Adenoviridae* (mastadenovirus), *Astroviridae* (mamastrovirus), *Hepeviridae* (orthohepevirus), *Parvoviridae* (bocaparvovirus), and *Picornaviridae* (Parechovirus, enterovirus, hepatovirus) families.



Figure 4. 3. Proportion of enteric viruses detected in wastewater samples collected during sampling year 1. Human viruses are annotated at the genus taxonomic level.

4. Discussion

4.1. Calicivirus Quantification in Wastewater and Clinical Presence

Quantification of NoV GII and SaV was carried out on wastewater samples to determine the presence and burden of these caliciviruses within the community. Detection rates between SaV and NoV were comparable during both sampling years. NoV obtained positive detection rates

similar to previous studies (Campos et al., 2016; Teixeira et al., 2020). SaV detection rates in raw wastewater ranging between 12.4% - 100% have been noted (Di Bartolo et al., 2013; Fioretti et al., 2016; Kaas et al., 2016; Kitajima et al., 2018; Kiulia et al., 2010; Mancini et al., 2019; Murray et al., 2013) with a positive detection rate of 92% observed in a recent study conducted in Arizona, U.S. (Kitajima et al., 2018). This is in agreement with the detection rates seen here demonstrating possible high prevalence of SaVs in the environment and circulation in human populations throughout the U.S. Concentrations of SaV were significantly greater than NoV GII during both sampling periods. Kaas et al., 2016 observed a similar trend with mean concentrations of SaV at 7.9×10^6 gc/L as compared to NoV GII 1.09×10^6 gc/L, though both viruses were detected in 100% of untreated wastewater samples. In contrast, a recent study investigating seasonal patterns of enteric viruses found SaV concentrations to be generally lower than noroviruses though both viruses displayed similar seasonal trends (Farkas et al., 2018b).

Indeed NoV is the leading causes of AGE in the U.S. with GII being the responsible genogroup in most outbreaks and pandemics (Glass et al., 2009), results here suggest an underestimated burden of SaV infections within the surrounding community. NoV and SaV are genetically similar producing near identical clinical manifestations proving difficult to distinguish without testing. SaV cases are not routinely reported in public health databases and therefore are linked to reported GI cases in this study. To compare the burden of NoV and SaV within the community one should considered the severity of illness, incubation period, duration of viral shedding, and viral shedding load.

SaV disease symptoms tend to be milder than those produced by NoV with a similar incubation period of 12-48 h (Lee et al., 2013; Oka et al., 2015). Milder symptoms cause by SaV infection likely impacts the clinical presence of the virus as compared to norovirus infections. Like NoV, SaV infections result in post-symptomatic viral shredding, which can persist for several weeks. Prolonged shedding has been observed in those with compromised immune systems and the elderly in SaV and NoV infections (Glass et al., 2009; Oka et al., 2015). Such information can distort the immediate burden of NoV and SaV infections but can provide insight into other community health indicators. These include prevalence of HIV, other autoimmune infections, and identification of the outbreak population. Additionally, viral shedding concentrations play a key role in loads captured in wastewater. Previous studies have observed viral shedding loads in feces up to 10^{11} genomic copies/g stool for SaV (Oka et al., 2015) and 10^{10} copies/g stool for NoV (Lee et al., 2007). Likewise, previously reported detection of viral loads in raw wastewater reveal similar concentrations between SaV and NoV (Haramoto et al., 2018). Considering this, results here indicate a higher burden of SaV infections during the time of sampling, at least as compared to the dominant NoV GII.

Several studies have noted similarities between caliciviruses in wastewater and epidemiological patterns within the study region. Iwai et al., 2009 observed high concentrations of NoV GII in wastewater following several NoV GII outbreaks within the surrounding community (Iwai et al., 2009). Likewise, similar trends have been noted elsewhere (Farkas et al., 2018a; Hellmér et al., 2014; Kaas et al., 2016). The lack of correlation between NoV and SaV concentrations and disease cases for norovirus and GI in this study may indicate interference from wastewater inhibitors, rapid changes in viral concentrations due to the short incubation period, or the lack of specific clinical

illness. Rotavirus, adenovirus, astroviruses, among others including non-viral pathogen may contribute significantly to GI cases reported in healthcare settings (Arena et al., 2014; Cunliffe et al., 2010; Räsänen et al., 2010).

4.2. Genogroup Classification of NoV and SaV and Diversity of Enteric Viruses in Wastewater

A metagenomic analysis was conducted on wastewater samples collected during SY1 to identify NoV and SaV genogroups circulating in the environment. Although less sensitive to genotype level variations as compare to amplicon sequencing, untargeted whole-genome sequencing has demonstrated practicality in this area with enhanced concentration or metagenomic approaches (Strubbia et al., 2019). The benefit of untargeted sequencing is the capability of detecting a multitude of human pathogens and characterize novel viruses without prior knowledge of the sample's microbiome. While we were not able to identify genotypes for NoV in this study, metagenomics discovered genogroups I and II. Norovirus Typing Tool classified several SaV contigs as genotype GI.1. Further confirmation is needed to confirm the presence of NoV and SaV genotypes in wastewater samples. Still results here indicate that untargeted NGS and metagenomics with optimized virus enrichment techniques is a promising approach to broad classification of calicivirus genogroups.

As expected, NoV GII was the most frequently detected genogroup in metagenomic samples. It is well known that NoV GII is the predominant strain in norovirus-associated outbreaks and sporadic cases (Glass et al., 2009; Harada et al., 2009; Rajko-Nenow et al., 2013) and has a global prevalence in wastewater (Farkas et al., 2018b; Iwai et al., 2009; Ueki et al., 2005). Previous

studies have also observed NoV GII to be most commonly detected in wastewater during the winter months in comparison to GI, which is seen during warmer climate (Kamel et al., 2010; Nordgren et al., 2009). NoV and SaV are both known to have distinct seasonal distribution (Oka et al., 2015; Robilotti et al., 2015), but year around sampling is required to draw such conclusions from this study. SaV GI was the only genogroup classified in wastewater samples. A recent study detected SaV GI.1 in 83% of samples considered (Mancini et al., 2019). Even further, SaV GI was the prevalent genogroup in wastewater collected from several geographical regions including Brazil (Fioretti et al., 2016), the United States (Kitajima et al., 2018), Italy (Di Bartolo et al., 2013), South Africa (Murray et al., 2013), Tunisia (Varela et al., 2018), and Japan (Kitajima et al., 2011). Moreover, SaV GI.1 was the responsible genotype in patients with acute diarrhea and sporadic outbreaks of AGE (Kumthip et al., 2018; Sánchez et al., 2018).

Among other enteric viruses detected, adenoviruses and astroviruses are common causes of diarrheal illness alongside caliciviruses. Adenoviruses belong to the mastadenovirus genus and are abundant in wastewater with subgroups F and G being a causative agent of gastroenteritis, mainly in children (Ghebremedhin, 2014). Indeed, the relatively low detection of adenoviruses (4/18) was surprising since these viruses are abundant in wastewater and persistent in environment (Bofill-Mas et al., 2006; Elmahdy et al., 2019; Fong et al., 2010). This is likely due to the reduced sensitivity of NGS since our previous study detected adenoviruses in the same set of samples with a 100% positive detection rate using qPCR (McCall et al., 2020). Nonetheless, subgroups F and G were detected in positive samples with > 90% similarly to the reference gene (data not shown).

Astroviruses, belonging to the mamastrovirus genus, were detected in 83% of samples and was the second most abundance enteric virus next to hepatovirus. Astroviruses have been detected in raw

wastewater in previous studies using metagenomics or PCR techniques (Meleg et al., 2006; Ng et al., 2012). Astroviruses are a leading cause of gastroenteritis in children accounting for up to 9% of acute nonbacterial gastroenteritis in children globally (Bosch et al., 2014). Additionally, in a previous study, astroviruses were the second most frequently detected viruses next to norovirus in stool samples collected from adults with acute diarrhea (Arena et al., 2014). Results here suggest a significant circulation of astrovirus infections within the service community during sampling. Lastly, human bocavirus (HBoV), of the bocaparvovirus genus, and parechovirus (HPeV) were detected in 55% and 67% of samples, respectively, and are also associated with diarrheal illnesses (Guido et al., 2016; Olijve et al., 2018). However, further work is needed to confirm what role HBoVs play in GI-related cases as they are commonly present alongside known etiological agents like adenovirus, rotavirus, and norovirus (Guido et al., 2016). Parechovirus and bocavirus have been detected in raw sewage and stool and in various geographical locations (Cantalupo et al., 2011; Hamza et al., 2017; Iaconelli et al., 2016; Räsänen et al., 2010; Rikhotso et al., 2020; Victoria et al., 2009).

5. Conclusion

It is well known that caliciviruses are leading causes of AGE globally. Concentrations of NoV and SaV were measured in untreated wastewater samples using qPCR to assess the burden of both viruses on a large metropolitan region in the United States. Metagenomic analysis was used to characterize NoV and SaV genogroups and identify the presence of other enteric viruses causing similar symptoms. SaV concentrations were significantly greater than NoV GII concentration throughout the study period suggesting a higher burden of SaV infections in the community during the time of sampling although there was not a significant correlation between virus concentrations

and norovirus and gastrointestinal cases. NoV GII and SaV GI were prevalent in wastewater samples. Metagenomic analysis was able to detect the presence of other important enteric pathogens and even ones possibly contributing significantly to gastrointestinal illness cases in the service community. This study highlights the need for routine monitoring of sapovirus infections and demonstrates the usefulness of metagenomics for viral surveillance.

Acknowledgements

This study was funded by National Science Foundation Award #1752773. We thank Anil Gosine (Detroit Water and Sewerage Department) and Michael Jurban (Great Lakes Water Authority) for allowing access to the Detroit wastewater treatment utility and assisting with sampling. We thank the Research Technology Support Facility (Michigan State University) for assisting with sequencing and Professor Shinhan Shiu (Michigan State University) for assisting with bioinformatics analysis.

REFERENCES

REFERENCES

Andrews, S., 2010. FastQC. Babraham Bioinforma.

- Arena, C., Amoros, J.P., Vaillant, V., Ambert-Balay, K., Chikhi-Brachet, R., Silva, N.J. Da, Varesi, L., Arrighi, J., Souty, C., Blanchon, T., Falchi, A., Hanslik, T., 2014. Acute diarrhea in adults consulting a general practitioner in France during winter: Incidence, clinical characteristics, management and risk factors. BMC Infect. Dis. 14, 1–7.
- Bibby, K., Viau, E., Peccia, J., 2011. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. Lett. Appl. Microbiol.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., Schneider, M., 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 31, 365–370.
- Bofill-Mas, S., Albinana-Gimenez, N., Clemente-Casares, P., Hundesa, A., Rodriguez-Manzano, J., Allard, A., Calvo, M., Girones, R., 2006. Quantification and stability of human adenoviruses and polyomavirus JCPyV in wastewater matrices. Appl. Environ. Microbiol.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics.
- Bosch, A., Pintó, R.M., Guix, S., 2014. Human astroviruses. Clin. Microbiol. Rev. 27, 1048–1074.
- Campos, C.J.A., Avant, J., Lowther, J., Till, D., Lees, D.N., 2016. Human norovirus in untreated sewage and effluents from primary, secondary and tertiary treatment processes. Water Res. 103, 224–232.
- Cannon, J.L., Barclay, L., Collins, N.R., Wikswo, M.E., Castro, C.J., Magaña, C., Gregoricus, N., Marine, R.L., 2017. Genetic and Epidemiologic Trends of Norovirus Outbreaks in the United Viruses. J. Clin. Microbiol. 55, 2208–2221.
- Cantalupo, P.G., Calgua, B., Zhao, G., Hundesa, A., Wier, A.D., Katz, J.P., Grabe, M., Hendrix, R.W., Girones, R., Wang, D., Pipas, J.M., 2011. Raw sewage harbors diverse viral populations. MBio 2, 1–11.
- Cunliffe, N.A., Booth, J.A., Elliot, C., Lowe, S.J., Sopwith, W., Kitchin, N., Nakagomi, O., Nakagomi, T., Hart, C.A., Regan, M., 2010. Healthcare-associated viral gastroenteritis among children in a large pediatric hospital, United Kingdom. Emerg. Infect. Dis. 16, 55–62.
- Di Bartolo, I., Ponterio, E., Battistone, A., Bonomo, P., Cicala, A., Mercurio, P., Triassi, M., Pennino, F., Fiore, L., Ruggeri, F.M., 2013. Identification and Genotyping of Human Sapoviruses Collected from Sewage Water in Naples and Palermo, Italy, in 2011. Food

Environ. Virol. 5, 236–240.

- Elmahdy, E.M., Ahmed, N.I., Shaheen, M.N.F., Mohamed, E.C.B., Loutfy, S.A., 2019. Molecular detection of human adenovirus in urban wastewater in Egypt and among children suffering from acute gastroenteritis. J. Water Health.
- Farkas, K., Cooper, D.M., McDonald, J.E., Malham, S.K., de Rougemont, A., Jones, D.L., 2018a. Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine receiving waters. Sci. Total Environ. 634, 1174–1183.
- Farkas, K., Marshall, M., Cooper, D., McDonald, J.E., Malham, S.K., Peters, D.E., Maloney, J.D., Jones, D.L., 2018b. Seasonal and diurnal surveillance of treated and untreated wastewater for human enteric viruses. Environ. Sci. Pollut. Res. 25, 33391–33401.
- Fioretti, J.M., Rocha, M.S., Fumian, T.M., Ginuino, A., da Silva, T.P., de Assis, M.R., Rodrigues, J.D.S., Carvalho-Costa, F.A., Miagostovich, M.P., 2016. Occurrence of human sapoviruses in wastewater and stool samples in Rio De Janeiro, Brazil. J. Appl. Microbiol. 121, 855–862.
- Fong, T.T., Phanikumar, M.S., Xagoraraki, I., Rose, J.B., 2010. Quantitative detection of human adenoviruses in wastewater and combined sewer overflows influencing a Michigan river. Appl. Environ. Microbiol.
- Ghebremedhin, B., 2014. Human adenovirus: Viral pathogen with increasing importance. Eur. J. Microbiol. Immunol. 4, 26–33.
- Glass, R.I., Parashar, U.D., Estes, M.K., 2009. Norovirus gastroenteritis. N. Engl. J. Med.
- GLWA, 2018. Capital Improvement Plan 2019-2023.
- Guido, M., Tumolo, M.R., Verri, T., Romano, A., Serio, F., De Giorgi, M., De Donno, A., Bagordo, F., Zizza, A., 2016. Human bocavirus: Current knowledge and future challenges. World J. Gastroenterol. 22, 8684–8697.
- Hall, A.J., Lopman, B.A., Payne, D.C., Patel, M.M., Gastañaduy, P.A., Vinjé, J., Parashar, U.D., 2013. Norovirus disease in the United States. Emerg. Infect. Dis. 19, 1198–1205.
- Hamza, H., Leifels, M., Wilhelm, M., Hamza, I.A., 2017. Relative Abundance of Human Bocaviruses in Urban Sewage in Greater Cairo, Egypt. Food Environ. Virol. 9, 304–313.
- Harada, S., Okada, M., Yahiro, S., Nishimura, K., Matsuo, S., Miyasaka, J., Nakashima, R., Shimada, Y., Ueno, T., Ikezawa, S., Shinozaki, K., Katayama, K., Wakita, T., Takeda, N., Oka, T., 2009. Surveillance of pathogens in outpatients with gastroenteritis and characterization of sapovirus strains between 2002 and 2007 in Kumamoto Prefecture, Japan. J. Med. Virol.

Haramoto, E., Kitajima, M., Hata, A., Torrey, J.R., Masago, Y., Sano, D., Katayama, H., 2018. A

review on recent progress in the detection methods and prevalence of human enteric viruses in water. Water Res. 135, 168–186.

- Hellmér, M., Paxéus, N., Magnius, L., Enache, L., Arnholm, B., Johansson, A., Bergström, T., Norder, H., 2014. Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. Appl. Environ. Microbiol. 80, 6771–6781.
- Iaconelli, M., Divizia, M., Della Libera, S., Di Bonito, P., La Rosa, G., 2016. Frequent Detection and Genetic Diversity of Human Bocavirus in Urban Sewage Samples. Food Environ. Virol.
- Iwai, M., Hasegawa, S., Obara, M., Nakamura, K., Horimoto, E., Takizawa, T., Kurata, T., Sogen, S.I., Shiraki, K., 2009. Continuous presence of noroviruses and sapoviruses in raw sewage reflects infections among inhabitants of Toyama, Japan (2006 to 2008). Appl. Environ. Microbiol. 75, 1264–1270.
- Jones, B., Cushingberry, G Jr., Ayers J., Benson, S., Castaneda-Lopez, R., Leland, G., Sheffield, M., A.S., 2015. Rehabilitation of the rectangular primary clarifiers, electrical/mechanical buildings and pipe gallery. Detroit.
- Jothikumar, N., Cromeans, T.L., Sobsey, M.D., Robertson, B.H., 2005. Development and evaluation of a broadly reactive TaqMan assay for rapid detection of hepatitis A virus. Appl. Environ. Microbiol. 71, 3359–3363.
- Kaas, L., Gourinat, A.C., Urbès, F., Langlet, J., 2016. A 1-Year Study on the Detection of Human Enteric Viruses in New Caledonia. Food Environ. Virol. 8, 46–56.
- Kamel, A.H., Ali, M.A., El-Nady, H.G., Aho, S., Pothier, P., Belliot, G., 2010. Evidence of the co-circulation of enteric viruses in sewage and in the population of Greater Cairo. J. Appl. Microbiol.
- Kitajima, M., Haramoto, E., Phanuwan, C., Katayama, H., 2011. Genotype distribution of human sapoviruses in wastewater in Japan. Appl. Environ. Microbiol. 77, 4226–4229.
- Kitajima, M., Rachmadi, A.T., Iker, B.C., Haramoto, E., Gerba, C.P., 2018. Temporal variations in genotype distribution of human sapoviruses and Aichi virus 1 in wastewater in Southern Arizona, United States. J. Appl. Microbiol. 124, 1324–1332.
- Kiulia, N.M., Netshikweta, R., Page, N.A., Van Zyl, W.B., Kiraithe, M.M., Nyachieo, A., Mwenda, J.M., Taylor, M.B., 2010. The detection of enteric viruses in selected urban and rural river water and sewage in Kenya, with special reference to rotaviruses. J. Appl. Microbiol. 109, 818–828.
- Kroneman, A., Vennema, H., Deforche, K., Avoort, H., Peñaranda, S., Oberste, M.S., Vinjé, J., Koopmans, M., 2011. An automated genotyping tool for enteroviruses and noroviruses. J. Clin. Virol. 51, 121–125.

- Kumthip, K., Khamrin, P., Maneekarn, N., 2018. Molecular epidemiology and genotype distributions of noroviruses and sapoviruses in Thailand 2000-2016: A review. J. Med. Virol.
- Le Guyader, F.S., Parnaudeau, S., Schaeffer, J., Bosch, A., Loisy, F., Pommepuy, M., Atmar, R.L., 2009. Detection and quantification of noroviruses in shellfish. Appl. Environ. Microbiol. 75, 618–624.
- Lee, N., Chan, M.C.W., Wong, B., Choi, K.W., Sin, W., Lui, G., Chan, P.K.S., Lai, R.W.M., Cockram, C.S., Sung, J.J.Y., Leung, W.K., 2007. Fecal viral concentration and diarrhea in norovirus gastroenteritis. Emerg. Infect. Dis. 13, 1399–1401.
- Lee, R.M., Lessler, J., Lee, R.A., Rudolph, K.E., Reich, N.G., Perl, T.M., Cummings, D.A.T., 2013. Incubation periods of viral gastroenteritis: A systematic review. BMC Infect. Dis. 13, 1.
- Mancini, P., Bonanno Ferraro, G., Iaconelli, M., Suffredini, E., Valdazo-González, B., Della Libera, S., Divizia, M., La Rosa, G., 2019. Molecular characterization of human Sapovirus in untreated sewage in Italy by amplicon-based Sanger and next-generation sequencing. J. Appl. Microbiol. 126, 324–331.
- McCall, C., Wu, H., Miyani, B. and Xagoraraki, I., 2020. Identification of multiple potential viral diseases in a large urban center using wastewater surveillance. Water Research, p.116160.
- MDHHS, 2018. Managing communicable diseases in schools [WWW Document]. URL https://www.michigan.gov/documents/mdch/Managing_CD_in_Schools_FINAL_469824_7 .PDF (accessed 4.28.20).
- MDHHS, 2020. Michigan Disease Surveillance System background [WWW Document]. URL https://www.michigan.gov/mdhhs/0,5885,7-339-71550_5104_31274-96814--,00.html (accessed 4.28.20).
- Meleg, E., Jakab, F., Kocsis, B., Bányai, K., Melegh, B., Szucs, G., 2006. Human astroviruses in raw sewage samples in Hungary. J. Appl. Microbiol.
- Murray, T.Y., Mans, J., Taylor, M.B., 2013. Human calicivirus diversity in wastewater in south africa. J. Appl. Microbiol. 114, 1843–1853.
- Ng, T.F.F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B.S., Wommack, K.E., Delwart, E., 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J. Virol. 86, 12161–12175.
- Nordgren, J., Matussek, A., Mattsson, A., Svensson, L., Lindgren, P.E., 2009. Prevalence of norovirus and factors influencing virus concentrations during one year in a full-scale wastewater treatment plant. Water Res. 43, 1117–1125.
- Oka, T., Katayama, K., Hansman, G.S., Kageyama, T., Ogawa, S., Wu, F.T., White, P.A., Takeda,

N., 2006. Detection of human sapovirus by real-time reverse transcription-polymerase chain reaction. J. Med. Virol.

- Oka, T., Wang, Q., Katayama, K., Saifb, L.J., 2015. Comprehensive review of human sapoviruses. Clin. Microbiol. Rev. 28, 32–53.
- Olijve, L., Jennings, L., Walls, T., 2018. Human Parechovirus : an Increasingly Recognized Cause of. Clin. Microbiol. Rev. 31, 1–17.
- R Core Team, 2019. R: A language and environment for statistical computing. Vienna, Austria.
- Rajko-Nenow, P., Waters, A., Keaveney, S., Flannery, J., Tuite, G., Coughlan, S., O'Flaherty, V., Doré, W., 2013. Norovirus genotypes present in oysters and in effluent from a wastewater treatment plant during the seasonal peak of infections in ireland in 2010. Appl. Environ. Microbiol. 79, 2578–2587.
- Räsänen, S., Lappalainen, S., Kaikkonen, S., Hämäläinen, M., Salminen, M., Vesikari, T., 2010. Mixed viral infections causing acute gastroenteritis in children in a waterborne outbreak. Epidemiol. Infect.
- Rikhotso, M.C., Khumela, R., Kabue, J.P., Traoré-Hoffman, A.N., Potgieter, N., 2020. Predominance of human bocavirus genotype 1 and 3 in outpatient children with diarrhea from rural communities in South Africa, 2017–2018. Pathogens 9, 2017–2018.

Robilotti, E., Deresinski, S., Pinsky, B.A., 2015. Norovirus. Clin. Microbiol. Rev. 28, 134–164.

- Sánchez, G.J., Mayta, H., Pajuelo, M.J., Neira, K., Xiaofang, L., Cabrera, L., Ballard, S.B., Crabtree, J.E., Kelleher, D., Cama, V., Bern, C., Oshitani, H., Gilman, R.H., Saito, M., 2018. Epidemiology of Sapovirus Infections in a Birth Cohort in Peru. Clin. Infect. Dis. 66, 1858– 1863.
- Strubbia, S., Schaeffer, J., Oude Munnink, B.B., Besnard, A., Phan, M.V.T., Nieuwenhuijse, D.F., de Graaf, M., Schapendonk, C.M.E., Wacrenier, C., Cotten, M., Koopmans, M.P.G., Le Guyader, F.S., 2019. Metavirome sequencing to evaluate norovirus diversity in sewage and related bioaccumulated oysters. Front. Microbiol.
- Teixeira, P., Costa, S., Brown, B., Silva, S., Rodrigues, R., Valério, E., 2020. Quantitative PCR detection of enteric viruses in wastewater and environmental water sources by the Lisbon municipality: A case study. Water (Switzerland).
- U.S. EPA, 2001. Manual of Methods for Virology (Chapter 14).
- Ueki, Y., Sano, D., Watanabe, T., Akiyama, K., Omura, T., 2005. Norovirus pathway in water environment estimated by genetic analysis of strains from patients of gastroenteritis, sewage , treated wastewater , river water and oysters 39, 4271–4280.
- van Beek, J., de Graaf, M., Al-Hello, H., Allen, D.J., Ambert-Balay, K., Botteldoorn, N., Brytting, M., Buesa, J., Cabrerizo, M., Chan, M., Cloak, F., Di Bartolo, I., Guix, S., Hewitt, J., Iritani, N., Jin, M., Johne, R., Lederer, I., Mans, J., Martella, V., Maunula, L., McAllister, G., Niendorf, S., Niesters, H.G., Podkolzin, A.T., Poljsak-Prijatelj, M., Rasmussen, L.D., Reuter, G., Tuite, G., Kroneman, A., Vennema, H., Koopmans, M.P.G., 2018. Molecular surveillance of norovirus, 2005–16: an epidemiological analysis of data collected from the NoroNet network. Lancet Infect. Dis. 18, 545–553.
- Varela, M.F., Ouardani, I., Kato, T., Kadoya, S., Aouni, M., Sano, D., Romalde, J.L., 2018. Sapovirus in wastewater treatment plants in Tunisia: Prevalence, removal, and genetic characterization. Appl. Environ. Microbiol.
- Victoria, J.G., Kapoor, A., Li, L., Blinkova, O., Slikas, B., Wang, C., Naeem, A., Zaidi, S., Delwart, E., 2009. Metagenomic Analyses of Viruses in Stool Samples from Children with Acute Flaccid Paralysis □ 83, 4642–4651.
- Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., Mardis, E.R., Hickenbotham, M., Magrini, V., Eldred, J., Latreille, J.P., Wilson, R.K., Ganem, D., DeRisi, J.L., 2003. Viral discovery and sequence recovery using DNA microarrays. PLoS Biol.

CONCLUSIONS AND SIGNIFICANCE

Wastewater-based epidemiology (WBE) may very well guide us into the next era of public health surveillance and transform epidemiology practices. Objectives of this study were to investigate human viruses in wastewater using viromics and molecular approaches for early detection and identification of viral diseases circulating in large human populations.

Chapter 1 discusses applications of viral metagenomics and reviews standard processes and best practices for viral discovery in metagenomic analysis. An intensive review was performed to determine approaches to virus concentration methods, sequencing tools, processes, and outcomes of recent viral metagenomic studies. From this we proposed a standard workflow for virus detection in metagenomes collected from water environments. This approach included the virus-absorption elution (VIRDEL) method for virus concentration and use of target-specific database for virus classification.

Chapter 2 seeks to employ WBE for early detection of hepatitis A outbreaks in urban communities using RT-qPCR and metagenomics. RT-qPCR captured hepatitis A virus (HAV) loads in wastewater during peak hepatitis A outbreak and sporadic case conditions. Hepatitis A cases were strongly correlated with viral concentrations in wastewater during peak outbreak conditions when adjusting for disease patterns. Spikes in HAV concentration in wastewater were followed by increases in the number of reported cases approximately 7 to 9 days after sampling. Moreover, the sensitivity of WBE to capture a rise in disease occurrence depends on the extent of cases present within the community. Despite strong correlations between clinical cases and HAV viral concentrations in wastewater, more frequent and rigorous environmental sampling is needed to fully understand HAV patterns in wastewater under various conditions. Lastly, metagenomics detected the presence of three viral hepatitis types in untreated wastewater samples, these were HAV, hepatitis E virus (HEV), and hepatitis C virus (HCV). This demonstrates that molecular and sequencing approaches can work together to identify various human viruses circulating in the community, better forecast disease outbreaks, and facilitate monitoring strategies for disease prevention.

In Chapter 3, untreated wastewater samples were examined using metagenomics and qPCR to evaluate WBE as a tool for public health monitoring and identification of viral threats circulating within a large community. Metagenomics detected the presence of enteric and non-enteric viruses that cause clinically important diseases that were reported within the study area during the sampling year. Furthermore, findings reveal evidence of re-emerging vector-borne viruses. Results presented in this study suggests that WBE has the potential to advance the area of disease outbreak mitigation and improve public health responses to large scale outbreaks and viral pandemics.

Chapter 4 investigates the burden of sapovirus (SaV) and norovirus (NoV) GII in a large metropolitan area in the United States using wastewater surveillance. NoV and SaV genogroups were characterized and samples were screened for other enteric viruses commonly causing acute gastroenteritis (AGE) using metagenomics. SaV concentrations were significantly greater than NoV GII concentration throughout the study period suggesting a higher burden of SaV infections in the community during the time of sampling. However, there was no significant correlation between virus concentrations and norovirus and gastrointestinal cases. NoV GII and SaV GI were prevalent in wastewater samples. Metagenomic analysis detected the presence of other important

enteric pathogens and even ones possibly contributing more significantly to gastrointestinal illness cases in the service community. This work highlights the need for routine monitoring of sapovirus infections and demonstrates the usefulness of metagenomics for viral surveillance.