**Quantitation of human enteric viruses as alternative indicators of fecal pollution to evaluate wastewater treatment processes**

Audrey Garcia, Tri Le, Paul Jankowski, Kadir Yanaç, Qiuyan Yuan and Miguel Uyaguari-Díaz\*

Department of Microbiology, Faculty of Science, University of Manitoba, Winnipeg, MB, Canada R3T 2N2

\* Corresponding Author:

Miguel Uyaguari-Díaz

45 Chancellors Cir, Winnipeg, Manitoba, R3T 2N2, Canada

Email address: Miguel.Uyaguari@umanitoba.ca

# **ABSTRACT**

We investigated the potential use and quantitation of human enteric viruses in municipal wastewater samples of Winnipeg (Manitoba, Canada) as alternative indicators of contamination and evaluated the processing stages of the wastewater treatment plant. During the fall 2019 and winter 2020 seasons, samples of raw sewage, activated sludge, effluents, and biosolids (sludge cake) from the North End Sewage Treatment Plant (NESTP), the largest wastewater treatment plant in the City of Winnipeg, were collected. DNA and RNA enteric viruses, as well as the *uidA* gene found in *Escherichia coli* were targeted in the samples collected from the NESTP. Total nucleic acids from each wastewater treatment sample were extracted using a commercial spin-column kit. Enteric viruses were quantitated in the extracted samples via quantitative PCR using TaqMan assays.

The average gene copies assessed in the raw sewage were not significantly different (p-values ranged between 0.0547 and 0.7986) than the average gene copies assessed in the effluents for Adenovirus and CrAssphage (DNA viruses), Pepper Mild Mottle Virus (RNA virus), and *uidA* in terms of both volume and biomass. A significant reduction of these enteric viruses was observed consistently in activated sludge samples compared with those for raw sewage. Corresponding reductions in gene copies per volume and gene copies per biomass were also seen for *uidA* but were not statistically significant (p-value = 0.8769 and p-value = 0.6353, respectively). The higher gene copy numbers of enteric viruses and *E. coli* observed in the effluents may be associated with the 12-hour hydraulic retention time in the facility. Enteric viruses were found in gene copy numbers at least one order of magnitude higher than the *E. coli* marker *uidA*. This may indicate these enteric viruses can survive the wastewater treatment process and therefore suggest that viral-like particles are being released into the aquatic environment. Our results suggest Adenovirus, CrAssphage, and Pepper mild mottle virus can be used as complementary viral indicators of human fecal pollution.

# **INTRODUCTION**

The human fecal waste present in raw sewage (RS) contains pathogens that can cause numerous diseases. This can have a huge negative impact to public and aquatic health and the economy (Stachler, et al., 2017). Wastewater treatment plants (WWTPs) serve as protective barriers between communities and the environment by reducing the organic matter present in wastewater. Water quality is currently assessed using traditional markers such as coliforms and *Escherichia coli*, leaving other microbes such as viruses largely unexplored. The North End Sewage Treatment Plant (NESTP) in Winnipeg, Manitoba handles approximately 70% of the city’s wastewater treatment, serving over 400,000 people (City of Winnipeg, Water and Waste Department, 2020). The treatment process at the NESTP first involves RS undergoing primary treatment during which solids are removed. It then gets processed to the stage of activated sludge (AS), in which a heterotrophic cocktail of bacteria and protozoa degrades organic matter present in solid waste. After this treatment cycle, AS is removed. The treated water is disinfected and is discharged as effluents (EF) into the river (City of Winnipeg, Water and Waste Department, 2020). Approximately 200 million liters of EF are discharged per day (City of Winnipeg, Water and Waste Department, 2020).

The main indicator of biological contamination used in wastewater treatment screening is *E. coli*, a fecal coliform bacterium (Hood, Ness, & Blake, 1983). Itis present in the gut of humans and warm-blooded animals and widely used as the main indicator of fecal pollution including during the wastewater treatment process.Nevertheless, the use of only fecal bacteria indicators excludes other possible pathogen groups present, such as human enteric viruses. Targeting these viruses in EF could be an alternative method to monitor the wastewater treatment process. Within this context, in a study conducted by Dutilh et al. (2014), the DNA CrAssphage genome was targeted in a human fecal sample. With further bioinformatics testing, it was predicted that the CrAssphage genome is highly abundant, and it was identified in 73% of human fecal metagenomes surveyed (Dutilh, et al., 2014). In a study conducted by Zhang et al. (2006), the most abundant fecal virus they found in dry weight fecal matter was the plant RNA virus, Pepper mild mottle virus (PMMV).

In the present study, samples of RS, AS, EF, and biosolids/sludge cake (SC) from the NESTP were collected (during fall 2019 and winter 2020) to investigate the potential of quantitating human enteric viruses in wastewater samples as complementary indicators of contamination to evaluate the processing stages of wastewater treatment. DNA enteric viruses in this study include human Adenovirus (AdV) and CrAssphage, while RNA enteric viruses include PMMV, Noroviruses (NoV) of the genogroups GI and GII, Astrovirus (AstV), Sapovirus (SaV) and Rotavirus (RoV). We also studied the presence of a molecular marker for *E. coli*, the *uidA* gene, in the samples collected from the NESTP. An overview of the workflow is illustrated in *Fig. 1*.

# **MATERIALS AND METHODS**

**Sample Collection.** A liter RSASEF1 kgSCfrom the NESTP during fall 2019 and winter 2020. Each sample was 1-Lsterile polyethylene container Samples were collected on October 22nd, 2019 (Event 1) and November 28th, 2019 (Event 2) in the fall season. In the winter season, samples were collected on December 18th, 2019 (Event 3) and February 6th, 2020 (Event 4). SC samples were collected on the days of Events 3 and 4. These samples were kept at 4°C and processed within 24 hours of collection.

**Assessment of Ultrafiltration for Viral Recovery Efficiency.** Armored RNA (Asuragen, Inc., Austin, TX, USA), an artificial virus, was used to assess recovery efficiency of the ultrafiltration method here employed. We spiked in 40,000 copies of Armored RNA into 120 mL raw sewage samples collected in duplicates from the NESTP, but not part of this study. Primers (381F: 5’- AGCCTGTCAATACCTGCACC-3’ and 475R: 5’- CACGCTTAGATCTCCGTGCT-3’), and probe (420P: 5’ Cy5-AGAGTATGAGAGGTCGACGA-TAO 3’) were designed using Primer design tool of Geneious Prime version 2021.1.1 ([https://www.geneious.com](http://www.geneious.com/%22%20%5Ct%20%22_blank" \t "_blank)) and targeted a 95-bp region within the Armored RNA genome. This targeted 95-bp fragment was sent to Integrated DNA Technologies (IDT, Inc., Coralville, Iowa) to synthetize a gBlock construct. Serial dilutions of this synthetic fragment were used to generate standards and quantify GCNs of Armored RNA via reverse transcription quantitative PCR (RT-qPCR). Standards and raw sewage samples were run in triplicates.

Thermal cycling reactions were performed at 50 °C for 5 minutes, followed by 45 cycles of 95 °C for 10 and 60 °C for 30 seconds on a QuantStudio 5 Real-Time PCR System (Life Technologies, Carlsbad, CA, USA). Each 10-μl RT-qPCR mixture consisted of 10 μl of 2.5 µL of 4X TaqMan Fast Virus 1-Step Master Mix (Life Technologies, Carlsbad, CA, USA), 400 nM each primer, 200 nM probe, 2.5 μl of template and ultrapure DNAse/RNAse free distilled water (Promega Corporation, Fitchburg, WI, USA).

**Ultrafiltration of Wastewater Samples.** Each wastewater treatment sample (RS, AS, and EF), including Millipore Milli-Q water as a negative control, was first filtered via a funnel and cheesecloth to remove any solid waste or debris. 140 mL of each wastewater sample was then concentrated down using an ultrafiltration method with Centricon Plus-70 filter units (Millipore Corporation, Billerica, MA, USA). The steps of the ultrafiltration method were as follows. Using a sterile glass pipette, 70 mL of each wastewater sample was added into their correspondingly labeled sample filter cup pre-assembled with the filtrate collection cup. Each assembly was then sealed with a cap. The Centricon Plus-70 assemblies were placed into a swinging bucket rotor and centrifuged at 3000 x g for 30 minutes at 20°C. Subsequently, the filtrate was discarded, and the remaining 70 mL of the samples was added into their correspondingly labeled sample filter cup pre-assembled with the filtrate collection cup. Samples were spun down using the same speed and temperature for 45 minutes. After centrifugation, the sample filter cup was separated from the filtrate collection cup. The concentration collection cup was then turned upside down and placed on top of the sample filter cup. The device was carefully inverted and placed into the centrifuge. Centricon Plus-70 filter units were centrifuged at 800 x g for 2 minutes at 20°C. After this step, the concentrated sample was collected from the concentration cup via a micropipette. The final volume was measured for each wastewater sample. If needed, 10 mM Tris-HCl, pH 8.5 buffer (Qiagen Sciences, Maryland, MD) was added to the concentrate to make up a total volume of 250 μL. If the final volume of the concentrate was over 250 μL, Tris buffer was not added. Aliquots containing 250 μL were made and stored at 4°C and processed within 24h.

**Sludge Cake Preparation for Ultrafiltration.** To remove cells from the SC samples, a 1X phosphate-buffered solution (PBS) with 0.15M NaCl, 0.05% Tween-20, pH 7.5

was used. Approximately 30 g of SC sample per sampling event (Events 3 and 4) was collected and divided into six Falcon tubes for each event (~5-6 g per tube). Approximately 30 mL of PBS was added to each tube. The Falcon tubes filled with SC samples were homogenized by vortexing them at constant agitation for 15 minutes at 2500 rpm. These tubes were then centrifuged at a speed of 4500 x g for 50 minutes. The supernatant from each tube was subsequently recovered and transferred to a new sterile Falcon tube. For each sample event, 140 mL of supernatant was used for ultrafiltration as described previously.

**Nucleic Acid (DNA/RNA) Extraction and Fluorometric Assessment.** Once the final volume of concentrate was collected from each wastewater sample, we pretreated the sample with InhibitEX buffer (Qiagen Sciences, Maryland, MD) as indicated by the manufacturer. Then, QIAamp MinElute virus spin kit (Qiagen Sciences, Maryland, MD) was used to extract total nucleic acids from each wastewater sample. We followed the manufacturer’s instructions that included the use of Qiagen Protease and carrier RNA (Qiagen Sciences, Maryland, MD). Samples were eluted in 75 μL of Buffer AVE (Qiagen Sciences, Maryland, MD), quantified and stored at -80°C for downstream processes. We assessed nucleic acid concentration and purity using Qubit dsDNA high sensitivity and RNA assay kits in a Qubit 4 fluorometer (Invitrogen, Carlsbad, CA, USA). Qubit results can be found in Supplementary Materials (*Table S1*).

**qPCR Primers, Probes, and gBlocks Gene Fragments.** *Table 1* summarizes the primers and probes used in this study. Forward and reverse primers described in *Table 1* were used in the Primer-BLAST tool to extract gene target regions (Ye, et al., 2012). Extracted regions were then uploaded to the software Geneious to verify oligonucleotide sequences associated to the flanking regions and probe. The generated sequences were sent to Integrated DNA Technologies (IDT, Inc., Coralville, Iowa, USA) to generate the desired gBlocks constructs. IDT manufactured all the primers used for qPCR, as well as the probes Ast-P, Ring1a.2, and Ring 2.2 (*Table 1).* However, probes Sav124TP, Sav5TP, NSP3-P, AdV-P, PMMV-P, and CrAss-P were manufactured by Life Technologies (Carlsbad, CA, USA).



**Quantitative PCR Assays.** Taqman Environmental Master Mix 2.0 (Life Technologies, Carlsbad, CA, USA) was used for assays involving DNA enteric viruses and *uidA*, while 4x Taqman Fast Virus 1-Step Master Mix (Life Technologies, Carlsbad, CA, USA) was used for RNA enteric viruses. Each 10 μl qPCR reaction contained 500 nM of each forward primer and reverse primer and 250 nM of its designated probe when targeting both DNA and RNA viruses. Five μl of Environmental Master Mix was in each qPCR reaction for targeting DNA viruses, while 2.5 μl of 4x Fast Virus Master Mix was in each qPCR reaction for targeting RNA viruses. The *uidA* qPCR reaction consisted of 5 μl of Environmental Master Mix, 400 nM of each primer, 100 nM of probe. All qPCR reactions used 2 μl of template.

Each qPCR reaction was performed in triplicates on an ABI QuantStudio 5 PCR system (Applied Biosystems, Foster City, CA, USA). The DNA enteric viruses (AdV and CrAssphage) and *uidA* ran under the following conditions: 50.0°C for 2 minutes and 95.0°C for 10 minutes followed by 40 cycles of 95.0°C for 15 seconds and 60.0°C for 1 minute. The RNA enteric viruses (SaV, RoV, AstV, GI and GII NoV, PMMV) ran under the following conditions: 50.0°C for 5 minutes and 95.0°C for 20 seconds followed by 40 cycles of 95.0°C for 3 seconds and 60.0°C for 30 seconds. Raw qPCR output files can be found on GitHub (https://git.io/J8VJ6).

**Assessment of Gene Copy Numbers by Volume and Biomass.** Gene copy numbers (GCNs) were expressed in terms of sample (per mL or g of sample) and biomass (per ng of DNA or RNA). GCNs per mL of sample were calculated as previously described by Ritalahti et al. (2006). When calculating GCNs per mL of sample, the final volume recovered after filtering 140 mL of wastewater sample was used in the formula. For the SC samples, the mass of SC collected was used in the formula, giving results in units of GCNs per g of sample.

**Collection of Metadata for Sampling Events.** To perform Principal Component Analysis (PCA) and Spearman’s rank correlation analysis for EF samples, metadata pertinent to the sampling events was retrieved. Water quality parameters obtained from the NESTP were combined with their October 2019 monitoring data (City of Winnipeg, Water and Waste Department, 2019) to complete some of the missing fields. For each value found in neither document, data interpolation was performed by taking an average of the corresponding values for the days before and after the sampling event. In addition, we searched the Government of Canada’s historical weather database to obtain the mean temperature on the sampling dates and the total precipitation over three days before each sampling event (hereafter referred to as “precipitation”) (Environment and Climate Change Canada, 2021). The values for all parameters were transformed using log10, except for precipitation due to the presence of zero values. These variables were used with log10-transformed GCNs per mL sample for AdV, CrAssphage, PMMV, and *uidA* (targets with quantifiable qPCR readings for all replicates across all events) as input for downstream analyses (PCA and Spearman’s rank correlation analysis).

**Data Handling, Statistical Analysis, and Data Visualization.** Various applications were employed to process data at different steps of the pipeline. Input data, such as output from the qPCR instrument, was subjected to rudimentary formatting and cleaning in Microsoft Excel, which was then also used to calculate GCNs per mL or g sample and per ng nucleic acid.

R (R Core Team, 2021) and its integrated development environment RStudio (RStudio Team, 2021) were utilized to further process the data and perform statistical analyses and output visualizations. These operations included general linear models (and estimated pairwise differences) using the package *sasLM* version 0.6.0 (Bae, 2021), PCA (corresponding biplots were created using the package *ggbiplot* version 0.55 (Vu, 2011)), and Spearman’s correlation matrix using the package *Hmisc* version 4.5-0 (Harrell Jr., 2021). The package *reshape2* version 1.4.4 (Wickham, reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package, 2020) was used to reformat these correlation matrices to make them more compatible with other data-handling tools. Other packages utilized but not mentioned here will be further detailed in Supplementary Materials (*Table S2*). The R script used for analysis can be found on GitHub (https://git.io/J8VUl ).

Another software involved in data visualization was Tableau. Specifically, it was used to generate boxplots for GCNs per mL or g sample and per ng nucleic acid, as well as the heatmap representing the above-mentioned Spearman correlation matrix.

For all tests, a p-value of 0.05 was assumed to be the minimum level of significance.

# **RESULTS**

For our assessment of the ultrafiltration method used in this study, the recovery efficiency of Armored RNA as measured by RT-qPCR was estimated to be between 7.14% and 8.64% for both raw sewage samples.

The GCN values for the DNA and RNA viruses and *uidA* were transformed into log10 form. These values then ran a general linear model Tukey-Kramer analysis, and the means of each wastewater processing stage for each target were analyzed. The GCNs were expressed in terms of volume (mL) or weight (g) of sample and biomass (ng of nucleic acids). The presence of DNA and RNA viral gene copies and *uidA* in the Milli-Q water (negative control) samples across all Events 1-4 were negative. The orange-dotted lines in *Figs. 2-6* indicate the mean of the number of gene copies of each wastewater treatment sample across all events.

The average GCNs assessed in the RS were not significantly different (p-values ranged between 0.0547 and 0.7986) compared to the average GCNs assessed in the EF for the DNA enteric viruses (AdV and CrAssphage) (*Fig. 2*), PMMV (*Fig. 3*), and *uidA* (*E. coli*) (*Fig. 4*) in terms of both volume and biomass. GCNsof the DNA enteric viruses AS and consistently RS. Corresponding reductions in gene copies per volume and gene copies per biomass were also seen for *uidA*, although these reductions were not statistically significant, with p-values being 0.8769 for the former and 0.6353 for the latter. For all the aforementioned targets, there was a relatively higher number of gene copies observed in the EF across all events compared to AS samples.

NoV GI and GII were also targets for our study. Box plots of their GCNs across the different wastewater stages throughout Events 1-4 can be found in Supplementary Materials (*Fig. S1*). In Events 1 and 2 (Fall season), NoV GI was below qPCR detection limits for all samples (RS, AS, and EF). Also below detection limits were NoV GII GCNs for all samples collected in Event 2 and AS samples in Events 3 and 4 (Winter season). Among the quantifiable samples, statistically significant GCN differences in terms of volume/mass and biomass were calculated for the pairs of AS-EF (p-values were 4.483 x 10-6 and 3.226 x 10-7, respectively), AS-RS (1.658 x 10-6, 1.091 x 10-5), and AS-SC (1.481 x 10-9, 4.083 x 10-7). No other significant differences were detected among treatments for CGNs of NoV GI. There was not a significant difference between the mean gene copies of NoV GII in the RS and EF samples in terms of volume (p-value = 0.7377), but the difference was significant in terms of biomass (p-value = 0.0490). The corresponding quantities of all the other sample pairs were statistically significant when looking at both the volume/mass and biomass perspectives, with p-values ranging from 1.304 x 10-8 to 0.0046, except for AS-RS GCN difference in terms of biomass (p-value = 0.0637).

We looked at RoV gene copies across the various wastewater treatment stages from Event 1 to 4. The boxplots illustrating these results in terms of both sample and biomass can be found in the Supplementary Materials (*Fig. S2*). RoV GCNs were below detection limit for all samples collected in Events 1 and 2. Looking at the EF-SC pair, the mean GCNs differed significantly in terms of volume/mass (p-value = 2.649 x 10-7) but not biomass (p-value = 0.4298). No other significant GCN differences could be detected between RS and AS samples in terms of both volume (p-value = 0.415509) and biomass (p-value = 0.6662). The equivalent magnitudes for the remaining pairs per volume/mass and per biomass were statistically significant, with p-values being between 7.907 x 10-10 and 0.0243. In the present study, there was no detection of gene copies for AstV and SaV (Sav1, Sav124, and Sav5) in any of the wastewater samples across all events.

To eliminate the possibility of inhibitors or contaminants such as humic acids, additional qPCR tests using bovine serum albumin (data not shown) were conducted with environmental samples (including AS). No significant differences were observed between samples with and without the enzyme.

To investigate any potential relationship between collected data for EF samples, PCA was performed with log10-transformed variables. We found that three components (PC1, PC2, and PC3) explained 99.14% of the variance between variables. A summary of the weight of components is included in the Supplementary Materials (*Table S3*). PC1 and PC2 were used to create the biplot in *Fig. 5*. Biplots for PC1 versus PC3 and PC2 versus PC3 are included in the Supplementary Materials (*Figs. S3 and S4*).

Overall, based on the biplot of PC1 and PC2, samples from the four events were distinct from one another, as point clusters of the four events can be seen occupying different quadrants. PC1, explaining 54.9% of the observed variance, received a notable and positive contribution from COD, cBOD, BOD, and TOC. Strongly negative contributors to PC1 were mean temperature, grab filtered UVT, NOx-N, and TS. These observations were supported by subsequent Spearman’s rank correlation analysis (*Fig. 6*), as COD, cBOD, BOD, and TOC demonstrated strongly positive correlations with one another (rho ranging between 0.800 and 0.949) (p-value < 0.005) and strongly negative correlations with mean temperature, grab filtered UVT, NOx-N, and TS (rho ranging between -0.949 and -0.800) (p-value < 0.005). PC2 explained 31.9% of the variance between sampling events and showed a strong contribution from CrAssphage, *uidA*, and grab temperature. This observation was also supported by the Spearman’s rank correlation analysis showing these variables having strongly positive correlation with one another (rho ranging between 0.717 and 0.922) (p-value < 0.01). Additionally, in the biplot, the axes representing *E. coli* and fecal coliform only pointed towards the same quadrant, something reflected in their moderately positive Spearman’s coefficient (0.632) (p-value = 0.0273). However, it is worth noting that *uidA* and *E. coli* exhibited a moderately weak negative correlation (rho = -0.307), although it was not statistically significant (p-value = 0.3313). The two parameters with the strongest contribution against PC2 were grab pH and turbidity, which strongly correlated with each other as illustrated in the Spearman’s coefficient heatmap (rho = 0.800) (p-value = 0.0018).

# **DISCUSSION**

The ultrafiltration method used in this study was assessed and the recovery efficiency was estimated to be between 7.14% and 8.64%. This range was lower compared to other methods to concentrate viral particles such as JumboSep (13.38% ± 9.11%) or skimmed milk flocculation (15.27% ± 3.32%) spiked-in wastewater samples and using Armored RNA as internal control (Yanaç and Uyaguari, unpublished results). Viral particles may have been sorbed to biosolids present in wastewater samples. In this context, matrix has a significant effect for recovery of viral particles. When compared to other environmental matrices such as surface water samples, recovery efficiency is higher using ultrafiltration (tangential flow filtration) (32.6%±11.81%) and skimmed milk flocculation (42.64%± 15.12%) (Francis and Uyaguari, unpublished results). Samples with high content of particles or suspended solids tend to saturate filters and impact the recovery of viral particles (Aslan, Xagoraraki, Simmons, Rose, & Dorevitch, 2011; Karim, Rhodes, Brinkman, Wymer, & Fout, 2009; Uyaguari-Diaz, et al., 2016).

The GCNs were expressed in terms of biomass and volume (except for SC, which was expressed in gs of sample). The higher abundance and more stable signal over time of GCNs of AdV, CrAssphage, and PMMV relative to the results for our other assays make these target more representative targets for conducting comparisons with *E. coli*. This persistent presence is consistent with various longitudinal studies previously performed (Ballesté, et al., 2019; Farkas, et al., 2018; Farkas, et al., 2019; Hamza, Rizk, Gad, & Hamza, 2019; Nour, et al., 2021; Schmitz, Kitajima, Campillo, Gerba, & Pepper, 2016; Tandukar, Sherchan, & Haramoto, 2020; Worley‐Morse, Mann, Khunjar, Olabode, & Gonzalez, 2019).

A reduction of AdV, CrAssphage, PMMV and *uidA* GCNs was observed consistently in AS samples. This could be a result of viral particles being sorbed to larger fractions of organic matter that had been filtered by cheesecloth early in the sample-handling process or retained in the filtration devices as previously described. It is important to mention in each sampling event, samples were collected within a 2-hour period from RS 🡪 AS 🡪 EF consecutively. The higher GCNs of viruses and *E. coli* observed in the EF may be associated with the hydraulic retention time (HRT) (12 hours) in the facility and may not reflect wastewater treatment profiles at the time of collection. Other variables to consider are the overflow of sewage from rainy events and fluctuations in mixed liquor-suspended solids (Pérez, Guerrero, Orellana, Figuerola, & Erijman, 2019). In our study, there were 4.6 mm of precipitation for Event 1, which may have affected the results here obtained. In the PCA analysis (*Fig. 5*), the vector for precipitation points sharply towards data points representing Event 1, indicating a possible relationship. Precipitation was also found to have positive correlations with grab flow (rho = 0.775) and raw flow (rho = 0.775) (*Fig. 6*). Nonetheless, further studies are needed to corroborate the potential link between precipitation and microbial counts.

Moreover, the duration of anaerobic sludge digestion is 15 days (City of Winnipeg, Water and Waste Department, 2020). In this context, GCNs of *uidA* in the SC were significantly reduced by anaerobic digestion. This may explain why the gene copies of *uidA* in terms of biomass were lower in SC compared to all treatments (p-value < 0.02). The average gene copies across all wastewater stages (RS, AS, and EF) for *uidA* were not significantly different in terms of both volume and biomass. When compared to *uidA*, enteric viruses were found to be at least one order of magnitude more abundant than the *E. coli* marker.

GCNs of CrAssphage in terms of biomass in SC were significantly higher than RS (p-value = 0.004028) and AS (p-value = 5.877 x 10-5). For PMMV, SC samples had significantly more GCNs in terms of biomass than samples from other parts of the wastewater treatment process (p-values ranged from 1.487 x 10-5 to 0.03775). Since SC is the by-product of RS and AS using anaerobic digestion, this may mean that the presence of CrAssphage and PMMV may have been lower in the wastewater being treated in the AS, but higher in the solids. However, GCNs of AdV in terms of biomass were not significantly different between the AS and SC samples. Meanwhile, plant viruses such as PMMV remain more stable (in terms of biomass) during these digestion processes (Jumat, et al., 2017).

The higher presence of RoV gene copies in the EF (p-value = 0.0005918 in terms of sample and p-value = 0.001498 in terms of biomass) during the winter season may indicate a higher risk of transmission during cold seasons (Atabakhsh et al. 2020), since it has been previously reported that a greater presence of RoV in EF are found during the winter season (Li, et al., 2011).

The negative results of SaV (Sav1, Sav124, and Sav5) across all wastewater treatment stages during the fall and winter season are consistent with a report by Varela et al. (2018) using samples from a wastewater treatment plant in Tunisia. Their results did not support the general belief that the peak of detection of SaV occurs during the cold and rainy months of the year. However, quantitative detection of SaV in wastewater and river water in Japan showed an increased concentration of SaV in influents between winter and spring (December to May), but a decrease in SaV concentration during the summer and autumn months (July to October) (Haramoto, Katayama, Phanuwan, & Ohgaki, 2008). Yet another pattern of SaV presence was reported in France in 2011, as Sima et al. found the virus to be readily detected in influents but had no clear variations in numbers over the 9-month (October to June) duration of the study. Similarly, seasonal differences in SaV concentrations were not statistically significant in a 3-year study conducted by Song et al. in China between 2017 and 2019. As a result, there might be other factors that can influence wastewater SaV concentrations. For example, it has been speculated that pI could affect how viruses and their different strains behave in bioreactors (Miura, Okabe, Nakahara, & Sano, 2015).

The gene copies of NoV GI and GII were below the detection limit in many of the AS samples (in terms of both volume and biomass), but still relatively abundant in the EF. A possible explanation for the greatly reduced viral GCNs in AS samples is the high efficiency with which NoV GI and GII are removed, a notion supported by literature (Ibrahim, Hammami, Khelifi, Pothier, & Hassen, 2020; Kitajima, Iker, Pepper, & Gerba, 2014; Schmitz, Kitajima, Campillo, Gerba, & Pepper, 2016). Furthermore, taking into account the observation that these viruses were found in abundance in SC samples, another contributing factor could be limitations in the sample collection process, which might not have adequately retrieved the slurry part of the sludge where the viruses may be found in greater numbers as they might have sorbed to the larger fractions of the sludge solids. The relative abundance of NoV GI and GII gene copies in RS and EF during the winter months (December and February) and the absence of NoV GII in RS in October may be due to seasonal variability (Pérez, Guerrero, Orellana, Figuerola, & Erijman, 2019). However, the presence of NoV GI and GII gene copies in RS in Event 4 (February) is consistent with a study conducted by Flannery et al. (2012), in which the concentration of NoV GI and GII gene copies in the influents of a wastewater treatment plant were significantly higher during the winter months (January to March). This seasonal trend is also reflected colloquially through its sobriquet, the winter vomiting bug (Farkas, et al., 2021).

In a study conducted by El-Senousy, et al. in 2007, high numbers of AstV gene copies (per liter) in sewage water samples (from the Greater Cairo area in Egypt) were observed at the end of autumn and during the winter months, but the AstV concentrations tended to decrease as temperatures increased. These results are significantly different from the quantitation of AstV gene copies in the wastewater samples from the NESTP. There was no detection of AstV gene copies in all the wastewater treatment stages across all events. These results may be due to seasonal variability (Pérez, Guerrero, Orellana, Figuerola, & Erijman, 2019) and efficiency of the wastewater facility in removing this type of virus during the process, as discussed above.

Additionally, PCA and Spearman’s rank correlation analysis were conducted for EF samples to investigate potential connections between various physical, chemical, and biological parameters. PCA revealed that samples collected in different events from October to February were distinct from one another. This could indicate a seasonal variation in wastewater, at least in effluents. The hypothesis is consistent with previous literature (Comber, Gardner, & Ellor, 2020). Organic chemical parameters such as COD, cBOD, BOD, and TOC were notable positive contributors to PC1, while mean temperature, grab filtered UVT, NOx-N, and TS most negatively contributed to PC1. These observations were validated by subsequent Spearman’s rank correlation analysis showing statistically significant coefficients. Grab filtered UVT being inversely correlated with COD, cBOD, BOD, and TOC is consistent with the widespread use of UV radiation to regulate microbial growth in a variety of medium, including water (Raeiszadeh & Adeli, 2020). While it is quite possible that the naked enteric viruses here studied may have survived the wastewater treatment process, the viral GCNs quantified in the EF may also represent an overestimation of the actual number of infectious viral particles since qPCR detects both infective and non-infective agents and UV treatment influenced viral viability (Lizasoain, et al., 2017). Therefore, the interpretation of these results must be taken with cautiousness.

# **CONCLUSION**

Our study’s primary goal is to identify human enteric viruses with the potential to become alternative indicators of fecal pollution. Towards that end, we propose AdV, CrAssphage, and PMMV as more stable viral indicators of water quality due to their quantifiability illustrated in this investigation and past ones supporting our results, as detailed in the Discussion section. Regular monitoring of these organisms can be useful complements to current methods for assessing wastewater treatment processes. Such vigilance could be a helpful tool to assist public health efforts in the event of a viral outbreak.

Additionally, our study indicated that enteric viruses may have survived the wastewater treatment process and viral-like particles are possibly being released into the aquatic environment. Therefore, in addition to using such methods as UV radiation (which is currently used and shown in our study to be inversely correlated with biological parameters), we also suggest that WWTPs consider implementing technologies specifically designed to reduce the number of viral particles in all stages of the wastewater treatment process.

# **ABBREVIATIONS**

**AdV:** Adenovirus

**AS:** activated sludge

**AstV:** Astrovirus

**BOD:** biochemical oxygen demand

**cBOD:** carbonaceous biochemical oxygen demand

**COD:** chemical oxygen demand

**EF:** effluents

**GCN:** gene copy number

**HRT:** hydraulic retention time

**NESTP:** North End Sewage Treatment Plant

**NH4-N:** ammonium-nitrogen

**NoV:** Norovirus

**NOx-N:** nitrogen oxides - nitrogen

**PCA:** Principal Component Analysis

**PMMV:** Pepper mild mottle virus

**PO4-P:** orthophosphate as phosphorus

**RoV:** Rotavirus

**RS:** raw sewage

**RT-qPCR:** reverse transcription quantitative PCR

**SaV:** Sapovirus

**SC:** sludge cake

**sCOD:** soluble chemical oxygen demand

**TN:** total nitrogen

**TOC:** total organic carbon

**TP:** total phosphorus

**TS:** total solids

**TSS:** total suspended solids

***uidA*:** encodes the beta-glucuronidase enzyme

**WWTP:** wastewater treatment plant

# **AUTHOR CONTRIBUTIONS**

[How to organize – each author does what, or what was done by which author(s)?]

# **ACKNOWLEDGEMENTS**

Special thanks to the City of Winnipeg and Palwinder Singh, graduate student, Department of Civil Engineering at the University of Manitoba for sample collection. Research start-up funds grant No. 322388 were assigned to Dr. Miguel Uyaguari-Diaz at the University of Manitoba. We acknowledge The Faculty of Science, University of Manitoba, collaborative grant No. 52622 (Drs. Uyaguari and Yuan). We also thank Dr. Brian Mark and his staff members (University of Manitoba) for the use of laboratory equipment.

# **REFERENCES**

Bae, K.-S. (2021). sasLM: 'SAS' Linear Model. Retrieved from https://CRAN.R-project.org/package=sasLM

Ballesté, E., Pascual-Benito, M., Martín-Díaz, J., Blanch, A. R., Lucena, F., Muniesa, M., . . . García-Aljaro, C. (2019, May 15). Dynamics of crAssphage as a human source tracking marker in potentially faecally polluted environments. *Water Research, 155*, 233-244. doi:10.1016/j.watres.2019.02.042

Barrios-Hernández, M. L., Pronk, M., Garcia, H., Boersma, A., Brdjanovic, D., & Mark C.M. van Loosdrecht, C. M. (2020, 1 1). Removal of bacterial and viral indicator organisms in full-scale aerobic granular sludge and conventional activated sludge systems. *Water Research X, 6*, 100040. doi:10.1016/j.wroa.2019.100040

City of Winnipeg, Water and Waste Department. (2019, October). North End Water Pollution Control Centre Monitoring Data. Winnipeg, Manitoba, Canada. Retrieved July 21, 2021, from https://www.winnipeg.ca/waterandwaste/pdfs/sewage/ComplianceReporting/2019/oct/newpcc.pdf

City of Winnipeg, Water and Waste Department. (2020, October 8). *Sewage Treatment Plants*. Retrieved July 21, 2021, from City of Winnipeg: https://www.winnipeg.ca/waterandwaste/sewage/treatmentPlant/default.stm#tab-north-end-sewage-treatment-plant

Comber, S. D., Gardner, M. J., & Ellor, B. (2020, September). Seasonal variation of contaminant concentrations in wastewater treatment works effluents and river waters. *Environmental Technology, 41*(21), 2716-2730. doi:10.1080/09593330.2019.1579872

Dutilh, B. E., Cassman, N., McNair, K., Sanchez, S. E., Silva, G. G., Boling, L., . . . Edwards, R. A. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nature Communications, 5*(4498), 1-11. doi:10.1038/ncomms5498

ElNaker, N. A., Yousef, A. F., & Hasan, S. W. (2018, August). Effect of hydraulic retention time on microbial community structure in wastewater treatment electro-bioreactors. *MicrobiologyOpen, 7*(4). doi:10.1002/mbo3.590

El-Senousy, W. M., Guix, S., Abid, I., Pintó, R. M., & Bosch, A. (2007, January). Removal of astrovirus from water and sewage treatment plants, evaluated by a competitive reverse transcription-PCR. *Applied and Environmental Microbiology, 73*(1), 164-7. doi:10.1128/AEM.01748-06

Environment and Climate Change Canada. (2021). *Historical Data*. Retrieved July 21, 2021, from Government of Canada: https://climate.weather.gc.ca/historical\_data/search\_historic\_data\_e.html

Farkas, K., Adriaenssens, E. M., Walker, D. I., McDonald, J. E., Malham, S. K., & Jones, D. L. (2019, June). Critical Evaluation of CrAssphage as a Molecular Marker for Human-Derived Wastewater Contamination in the Aquatic Environment. *Food and Environmental Virology, 11*(2), 113-119. doi:10.1007/s12560-019-09369-1

Farkas, K., Green, E., Rigby, D., Cross, P., Tyrrel, S., Malham, S. K., & Jones, D. L. (2021, May 27). Investigating awareness, fear and control associated with norovirus and other pathogens and pollutants using best–worst scaling. *Scientific Reports, 11*.

Farkas, K., Marshall, M., Cooper, D., McDonald, J. E., Malham, S. K., Peters, D. E., . . . Jones, D. L. (2018, November). Seasonal and diurnal surveillance of treated and untreated wastewater for human enteric viruses. *Environmental Science and Pollution Research, 25*(33), 33391-33401. doi:10.1007/s11356-018-3261-y

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2021). mvtnorm: Multivariate Normal and t Distributions. Retrieved from http://CRAN.R-project.org/package=mvtnorm

Hamza, H., Rizk, N. M., Gad, M. A., & Hamza, I. A. (2019, November). Pepper mild mottle virus in wastewater in Egypt: a potential indicator of wastewater pollution and the efficiency of the treatment process. *Archives of Virology, 164*(11), 2707-2713. doi:10.1007/s00705-019-04383-x

Haramoto, E., Katayama, H., Phanuwan, C., & Ohgaki, S. (2008, March). Quantitative detection of sapoviruses in wastewater and river water in Japan. *Letters in Applied Microbiology, 46*(3), 408-13. doi:10.1111/j.1472-765X.2008.02330.x

Harrell Jr., F. E. (2021). Hmisc: Harrell Miscellaneous. Retrieved from https://CRAN.R-project.org/package=Hmisc

Hood, M. A., Ness, G. E., & Blake, N. J. (1983). Relationship among fecal coliforms, Escherichia coli, and Salmonella spp. in shellfish. *Applied and environmental microbiology, 45*(1), 122-126. doi:10.1128/aem.45.1.122-126.1983

Jumat, M. R., Hasan, N. A., Subramanian, P., Heberling, C., Colwell, R. R., & Hong, P.-Y. (2017). Membrane Bioreactor-Based Wastewater Treatment Plant in Saudi Arabia: Reduction of Viral Diversity, Load, and Infectious Capacity. *Water, 9*(7). doi:10.3390/w9070534

Kageyama, T., Kojima, S., Shinohara, M., Uchida, K., Fukushi, S., Hoshino, F. B., . . . Katayama, K. (2003, April). Broadly Reactive and Highly Sensitive Assay for Norwalk-Like Viruses Based on Real-Time Quantitative Reverse Transcription-PCR. *Journal of Clinical Microbiology, 41*(4), 1548-1557. doi:10.1128/JCM.41.4.1548-1557.2003

Li, D., Gu, A. Z., Zeng, S.-Y., Yang, W., He, M., & Shi, H.-C. (2011, May). Monitoring and evaluation of infectious rotaviruses in various wastewater effluents and receiving waters revealed correlation and seasonal pattern of occurrences. *Journal of Applied Microbiology, 110*(5), 1129-37. doi:10.1111/j.1365-2672.2011.04954.x

Liang, L., Goh, S. G., Vergara, G. G., Fang, H. M., Rezaeinejad, S., Chang, S. Y., . . . Gin, K. Y. (2015). Alternative Fecal Indicators and Their Empirical Relationships with Enteric Viruses, Salmonella enterica, and Pseudomonas aeruginosa in Surface Waters of a Tropical Urban Catchment. *Applied and Environmental Microbiology, 81*(3), 850-860. doi:10.1128/AEM.02670-14

Lizasoain, A., Tort, L., García, M., Gillman, L., Alberti, A., Leite, J., . . . Colina, R. (2017). Human enteric viruses in a wastewater treatment plant: evaluation of activated sludge combined with UV disinfection process reveals different removal performances for viruses with different features. *Letters in Applied Microbiology, 66*(3), 215-221. doi:DOI: 10.1111/lam.12839

Makowski, D., Ben-Shachar, M., Patil, I., & Lüdecke, D. (2020). Automated Results Reporting as a Practical Tool to Improve Reproducibility and Methodological Best Practices Adoption. Retrieved from https://github.com/easystats/report

Miura, T., Okabe, S., Nakahara, Y., & Sano, D. (2015, May 15). Removal properties of human enteric viruses in a pilot-scale membrane bioreactor (MBR) process. *Water Research, 75*, 282-291. doi:10.1016/j.watres.2015.02.046

Molecular Microbiology & Genomics Team, British Columbia Centre for Disease Control. (2017). Detecting Norovirus by Fast Real-Time RT-PCR. British Columbia, Canada.

Molecular Microbiology & Genomics Team, British Columbia Centre for Disease Control. (2017). Performing the GI Virus Panel by Real-Time PCR Procedure. British Columbia, Canada.

Nour, I., Hanif, A., Zakri, A. M., Al-Ashkar, I., Alhetheel, A., & Eifan, S. (2021, April 29). Human Adenovirus Molecular Characterization in Various Water Environments and Seasonal Impacts in Riyadh, Saudi Arabia. *International Journal of Environmental Research and Public Health, 18*(9), 4773. doi:10.3390/ijerph18094773

Oka, T., Katayama, K., Hansman, G. S., Kageyama, T., Ogawa, S., Wu, F.-T., . . . Takeda, N. (2006). Detection of human sapovirus by real-time reverse transcription-polymerase chain reaction. *Journal of Medical Virology, 78*(10), 1347-1353. doi:10.1002/jmv.20699

Pan, S., Tay, J.-H., He, Y.-X., & Tay, S. T.-L. (2004). The effect of hydraulic retention time on the stability of aerobically grown microbial granules. *Letters in Applied Microbiology, 38*(2), 158-163. doi:10.1111/j.1472-765x.2003.01479.x

Payment, P., Fortin, S., & Trudel, M. (1986, December). Elimination of human enteric viruses during conventional waste water treatment by activated sludge. *Canadian Journal of Microbiology, 32*(12), 922-5. doi:10.1139/m86-170

Pérez, M. V., Guerrero, L. D., Orellana, E., Figuerola, E. L., & Erijman, L. (2019, July 2). Time Series Genome-Centric Analysis Unveils Bacterial Response to Operational Disturbance in Activated Sludge. *mSystems, 4*(4). doi:10.1128/mSystems.00169-19

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Raeiszadeh, M., & Adeli, B. (2020, October 14). A Critical Review on Ultraviolet Disinfection Systems against COVID-19 Outbreak: Applicability, Validation, and Safety Considerations. *ACS Photonics*. doi:10.1021/acsphotonics.0c01245

Ritalahti, K. M., Amos, B. K., Sung, Y., Wu, Q., Koenigsberg, S. S., & Löffler, F. E. (2006). Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple Dehalococcoides strains. *Applied and Environmental Microbiology, 72*(4), 2765-74. doi:10.1128/AEM.72.4.2765-2774.2006

Rosario, K., Symonds, E. M., Sinigalliano, C., Stewart, J., & Breitbart, M. (2009). Pepper Mild Mottle Virus as an Indicator of Fecal Pollution. *Applied and Environmental Microbiology, 75*(22), 7261-7267. doi:10.1128/AEM.00410-09

RStudio Team. (2021). RStudio: Integrated Development Environment for R. Boston, MA: RStudio, PBC. Retrieved from http://www.rstudio.com/

Safferman, R., & Morris, M.-E. (1976). Assessment of virus removal by a multi-stage activated sludge process. *Water Research, 10*(5), 413-420. doi:10.1016/0043-1354(76)90059-2

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R.* New York: Springer.

Schmitz, B. W., Kitajima, M., Campillo, M. E., Gerba, C. P., & Pepper, I. L. (2016). Virus Reduction during Advanced Bardenpho and Conventional. *Environmental Science & Technology, 50*(17), 9524-9532. doi:10.1021/acs.est.6b01384

Sima, L. C., Schaeffer, J., Saux, J.-C. L., Parnaudeau, S., Elimelech, M., & Guyader, F. S. (2011, August). Calicivirus Removal in a Membrane Bioreactor Wastewater Treatment Plant. *Applied and Environmental Microbiology, 77*(15), 5170-5177. doi:10.1128/AEM.00583-11

Song, K., Lin, X., Liu, Y., Ji, F., Zhang, L., Chen, P., . . . Xu, A. (2021). Detection of Human Sapoviruses in Sewage in China by Next Generation Sequencing. *Food and Environmental Virology, 13*, 270-280. doi:10.1007/s12560-021-09469-x

Stachler, E., Akyon, B., Carvalho, N. A., Ference, C., & Bibby, K. (2018). Correlation of crAssphage qPCR Markers with Culturable and Molecular Indicators of Human Fecal Pollution in an Impacted Urban Watershed. *Environmental Science & Technology, 52*(13), 7505-7512. doi:10.1021/acs.est.8b00638

Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., & Shanks, O. C. (2017). Quantitative CrAssphage PCR Assays for Human Fecal Pollution Measurement. *Environmental Science & Technology, 51*(16), 9146–9154. doi:10.1021/acs.est.7b02703

Tandukar, S., Sherchan, S. P., & Haramoto, E. (2020). Applicability of crAssphage, pepper mild mottle virus, and tobacco mosaic virus as indicators of reduction of enteric viruses during wastewater treatment. *Scientific Reports, 10*(3616). doi:10.1038/s41598-020-60547-9

Therneau, T. (2021). A Package for Survival Analysis in R. Retrieved from https://CRAN.R-project.org/package=survival

Ushey, K., Allaire, J., Wickham, H., & Ritchie, G. (2020). rstudioapi: Safely Access the RStudio API. Retrieved from https://CRAN.R-project.org/package=rstudioapi

Varela, M. F., Ouardani, I., Kato, T., Kadoya, S., Aouni, M., Sano, D., & Romalde, J. L. (2018, March 1). Sapovirus in Wastewater Treatment Plants in Tunisia: Prevalence, Removal, and Genetic Characterization. *Applied and Environmental Microbiology, 84*(6). doi:10.1128/AEM.02093-17

Vu, V. Q. (2011). ggbiplot: A ggplot2 based biplot. Retrieved from http://github.com/vqv/ggbiplot

Wang, X. M., Wei, Z. M., Guo, J. M., Cai, J. M., Chang, H. M., Ge, Y. M., & Zeng, M. M. (2019, November). Norovirus Activity and Genotypes in Sporadic Acute Diarrhea in Children in Shanghai During 2014–2018. *The Pediatric Infectious Disease Journal, 38*(11), 1085-1089. doi:10.1097/INF.0000000000002456

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software, 40*(1), 1-29. doi:10.18637/jss.v040.i01

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* New York: Springer-Verlag.

Wickham, H. (2020, April 9). reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package. Retrieved from https://cran.r-project.org/package=reshape2

Wickham, H., & Bryan, J. (2019). readxl: Read Excel Files. Retrieved from https://CRAN.R-project.org/package=readxl

Wickham, H., & Bryan, J. (2021). usethis: Automate Package and Project Setup. Retrieved from https://CRAN.R-project.org/package=usethis

Wickham, H., & Seidel, D. (2020). scales: Scale Functions for Visualization. Retrieved from https://CRAN.R-project.org/package=scales

Wickham, H., François, R., Henry, L., & Müller, K. (2021). dplyr: A Grammar of Data Manipulation. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., Hester, J., & Chang, W. (2021). devtools: Tools to Make Developing R Packages Easier. Retrieved from https://CRAN.R-project.org/package=devtools

Worley‐Morse, T., Mann, M., Khunjar, W., Olabode, L., & Gonzalez, R. (2019, September). Evaluating the fate of bacterial indicators, viral indicators, and viruses in water resource recovery facilities. *Water Environment Research, 91*(9), 830-842. doi:10.1002/wer.1096

Zeileis, A., & Croissant, Y. (2010). Extended Model Formulas in R: Multiple Parts and Multiple Responses. *Journal of Statistical Software, 34*(1), 1-13. doi:10.18637/jss.v034.i01

Zeng, S. Q., Halkosalo, A., Salminen, M., Szakal, E. D., Puustinen, L., & Vesikari, T. (2008). One-step quantitative RT-PCR for the detection of rotavirus in acute gastroenteritis. *Journal of Virological Methods, 153*(2), 238-40. doi:10.1016/j.jviromet.2008.08.004

Zhang, T., Breitbart, M., Lee, W. H., Run, J.-Q., Wei, C. L., Soh, S. W., . . . Ruan, Y. (2006, January). RNA Viral Community in Human Feces: Prevalence of Plant Pathogenic Viruses. *PLOS Biology, 4*(1). doi:10.1371/journal.pbio.0040003