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

Audrey Garcia, Tri Le, Paul Jankowski 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* (*E. 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.

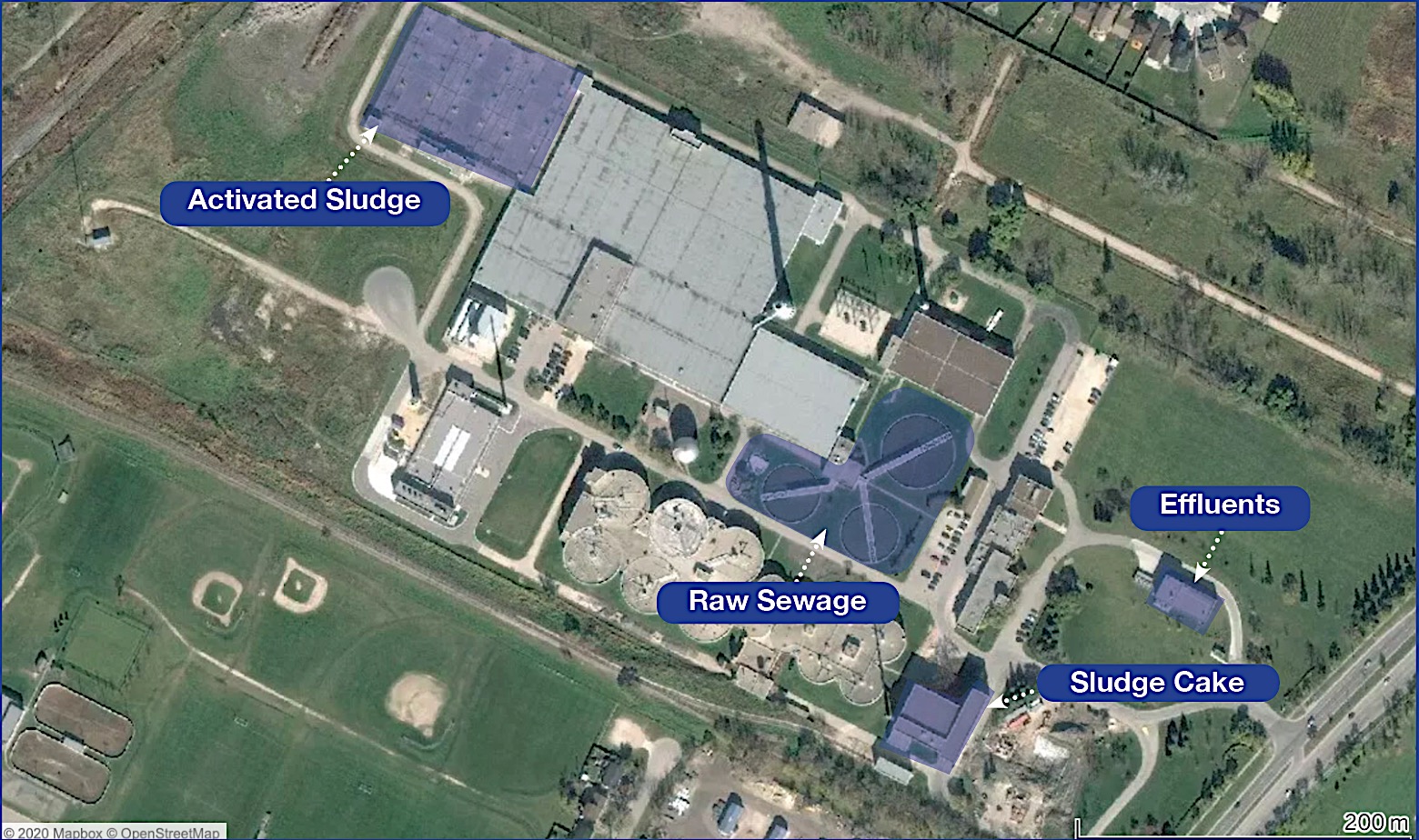
The average gene copies assessed in the raw sewage were lower but not significantly different (p-values ranged between 0.0547 and 0.7986) than the average gene copies assessed in the effluents for Adenovirus and CrAssphage (DNA viruses), Pepper Mild Mottle Virus (RNA virus), and *uidA* in terms of both 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 non-enveloped enteric viruses can survive the wastewater treatment process and therefore suggest that viral-like particles are being released into the aquatic environment.

**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 70% of the city’s wastewater treatment, serving over 400,000 people (City of Winnipeg, 2019a). 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, 2019a). Approximately 200 million liters of EF are discharged per day (City of Winnipeg, 2019b).

The main indicator of contamination used in wastewater treatment screening is *E. coli*, a fecal coliform bacterium (Hood et al., 1982 – [[aem00170-0140.pdf (nih.gov)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC242241/pdf/aem00170-0140.pdf)]). 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, 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 Adenovirus and CrAssphage, while RNA enteric viruses include Pepper Mild Mottle Virus (PMMV), Noroviruses GI & GII, Astrovirus, Sapovirus and Rotavirus. We also studied the presence of a molecular marker for *E. coli*, the *uidA* gene, in the samples collected from the NESTP.



**Figure 1.** Satellite photo of the North End Sewage Treatment Plant.

**MATERIALS AND METHODS**

**Sample Collection**

A liter RSASEF1 kgSCfrom the NESTP during fall 2019 and winter 2020. Each sample was 1LSamples 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 h of collection.

**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). 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, 1X phosphate-buffered solution (PBS) and 0.01% Tween pH 7.4 were 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 (as described above), 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 S number].

**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. 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) 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. However, probes Sav124TP, Sav5TP, NSP3-P, AdV-P, PMMV-P, and CrAss-P were manufactured by Life Technologies (Carlsbad, CA, USA).

**Table 1.** Primers and Probes used in the present study.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target** | **DNA or RNA** | **Primer/Probe** | **Sequence (5’-3’)** | **Genomic Target** | **References** |
| Sapovirus | RNA | Sav1F | TTG GCC CTC GCC ACC TAC | Junction of polymerase and capsid |  |
| Sav5F | TTT GAA CAA GCT GTG GCA TGC TAC |
| Sav124F | GAY CAS GCT CTC GCY ACC TAC |
| Sav124R | CCC TCC ATY TCA AAC ACT A |
| Sav124TPa (Probe) | FAM-CCR CCT ATR AAC CA |
| Sav5TPa (Probe) | FAM-TGC CAC CAA TGT ACC A |
| Rotavirus Type A | RNA | NSP3-F | ACC ATC TWC ACR TRA CCC TCT ATG AG | Non-structural Protein 3 |  |
| NSP3-R | GGT CAC ATA ACG CCC CTA TAG C |
| NSP3-Pa (Probe) | VIC-AGT TAA AAG CTA ACA CTG TCA AA |
| Adenovirus 40/41 | DNA | AdV-F | GCC TGG GGA ACA AGT TCA G | Hexon |  |
| AdV-R | ACG GCC AGC GTA AAG CG |
| AdV-Pa (Probe) | NED-ACC CAC GAT GTA ACC AC |
| Astrovirus | RNA | Ast-F | AAG CAG CTT CGT GAR TCT GG | Junction of polymerase and capsid |  |
| Ast-R | GCC ATC RCA CTT CTT TGG TCC |
| Ast-Pb (Probe) | Cy5-CAC AGA AGA GCA ACT CCA TCG CAT TTG |
| GI | RNA | Cog1F-flap | AATAAATCATAACGYTGGATGCGNTTYCATGA | Norovirus GI |  |
| Cog1R-flap | AATAAATCATAACTTAGACGCCATCATCATTYAC |
| Ring1a.2 (Probe) | 6-FAM- AGATYGCGR/ZEN/ TCYCCTGTCCA -IBFQ |
| GII | RNA | Cog2F-flap | AATAAATCATAACARGARBCNATGTTYAGRTGGAT GAG | Norovirus GII |  |
| Cog2R-flap | AATAAATCATAATCGACGCCATCTTCATTCACA |
|  |
| Ring 2.2 (Probe) | JOE - TGGGAGGGY/ZEN/ GATCGCAATCT - IBFQ |  |
| CrAssphage | DNA | CrAss-F | CAG AAG TAC AAA CTC CTA AAA AAC GTA GAG | Genomic base pair region: 14731 bp-14856 bp |  |  |
| CrAss-R | GAT GAC CAA TAA ACA AGC CAT TAG C |  |
| CrAss-P (Probe) | FAM-AAT AAC GAT TTA CGT GAT GTA AC |  |
| Pepper Mild Mottle Virus | RNA | PMMV-FP1-rev | GAG TGG TTT GAC CTT AAC GTT TGA | 1878 bp-1901 bpc  and |  |  |
| 1945 bp-1926 bpc |  |  |
| PMMV-RP1 | TTG TCG GTT GCA ATG CAA GT |  |  |  |
| PMMV-P (Probe) | FAM-CCT ACC GAA GCA AAT G |  |  |  |
| *Escherichia coli* | DNA | uidA\_784F | GTG TGA TAT CTA CCC GCT TCG C |  |  |  |
| uidA\_866R | AGA ACG GTT TGT GGT TAA TCA GGA | *uidA* |  |  |
| EC807d probe | FAM-TCGGCATCCGGTCAGTGGCAGT-BHQ1 |  |  |  |
|  | | aQuencher: MGB-NFQ. | | | |  |
| bQuencher: Tao-IBDRQ. | | | |  |
| cCorresponding nucleotide position of GenBank accession number M81413 (PMMoV strain S) | | | |  |
| dQuencher: Iowa Black fluorescent | | | |  |



**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, 0.4 μM of each primer, 0.1 μM of probe. All qPCR reactions used 2 μl of template.

Each qPCR reaction was performed in triplicates on the ABI QuantStudio 5 PCR system (Applied Biosystems, Foster City, CA, USA). The DNA enteric viruses (Adenovirus 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 (Sapovirus, Rotavirus, Astrovirus, GI and GII Norovirus, 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 [GitHub link].

**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 [need to make reference for this link - [2019-10 NEWPCC.xlsx (winnipeg.ca)](https://www.winnipeg.ca/waterandwaste/pdfs/sewage/ComplianceReporting/2019/oct/newpcc.pdf)] 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”) [[Historical Data - Climate - Environment and Climate Change Canada (weather.gc.ca)](https://climate.weather.gc.ca/historical_data/search_historic_data_e.html)]. 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 Adenovirus, 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 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* [Bae, 2021], PCA (corresponding biplots were created using the package *ggbiplot* [Vu, 2011]), and Spearman’s correlation matrix using the package *Hmisc* [Harrell et al., 2021]. The package *reshape2* [Wickham, 2007] was used to reformat these correlation matrices to make them more compatible with other data-handling tools. Additionally, of much assistance during the analysis process were several general-purpose and dependency packages, which will be further detailed in Supplementary Materials [Table S number]. The R script used for analysis can be found on GitHub ([GitHub link]).

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.

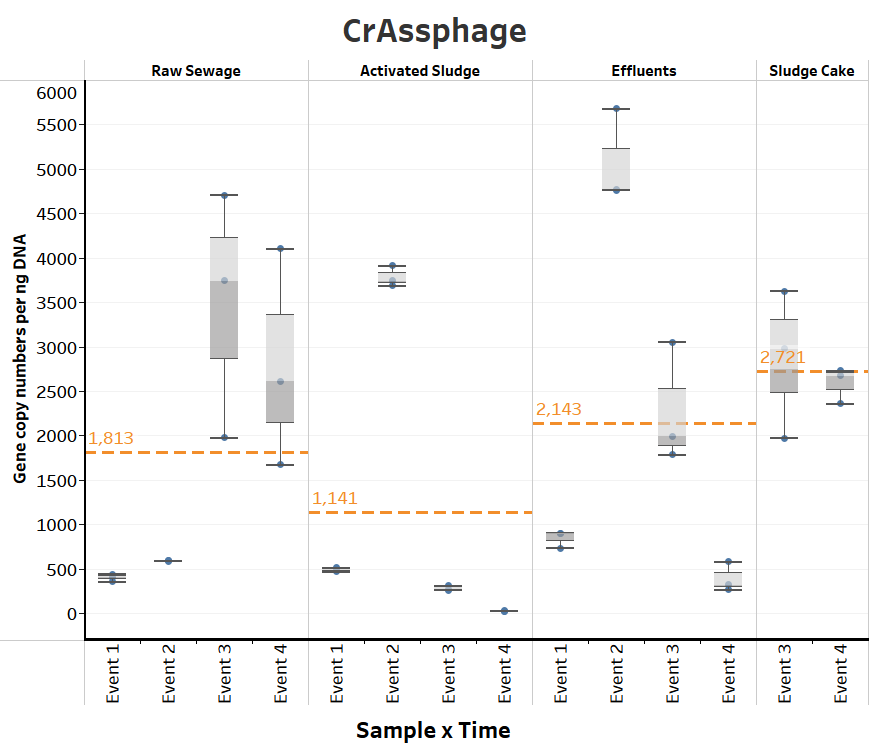
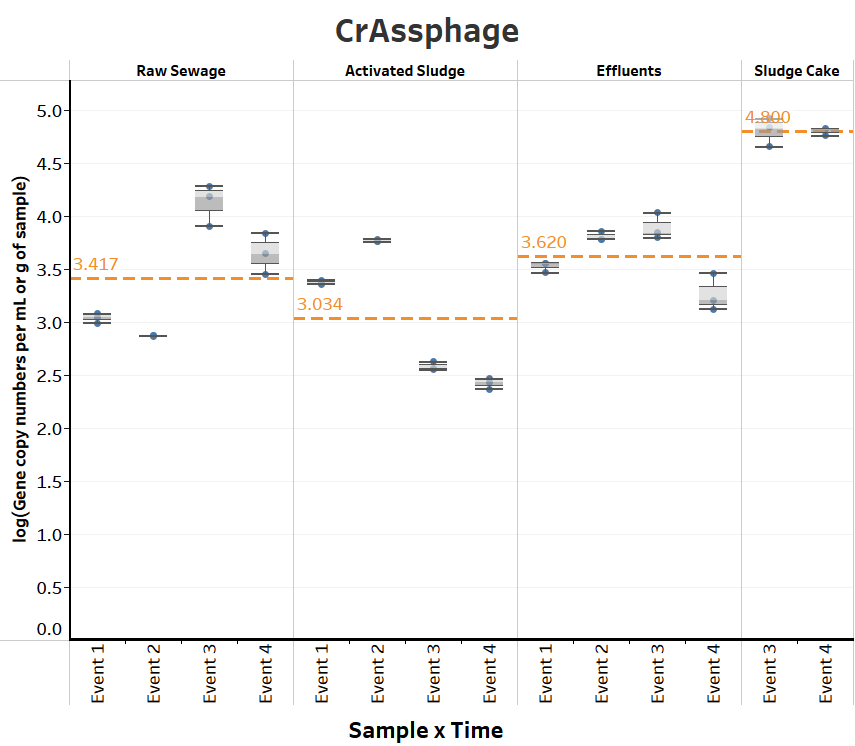
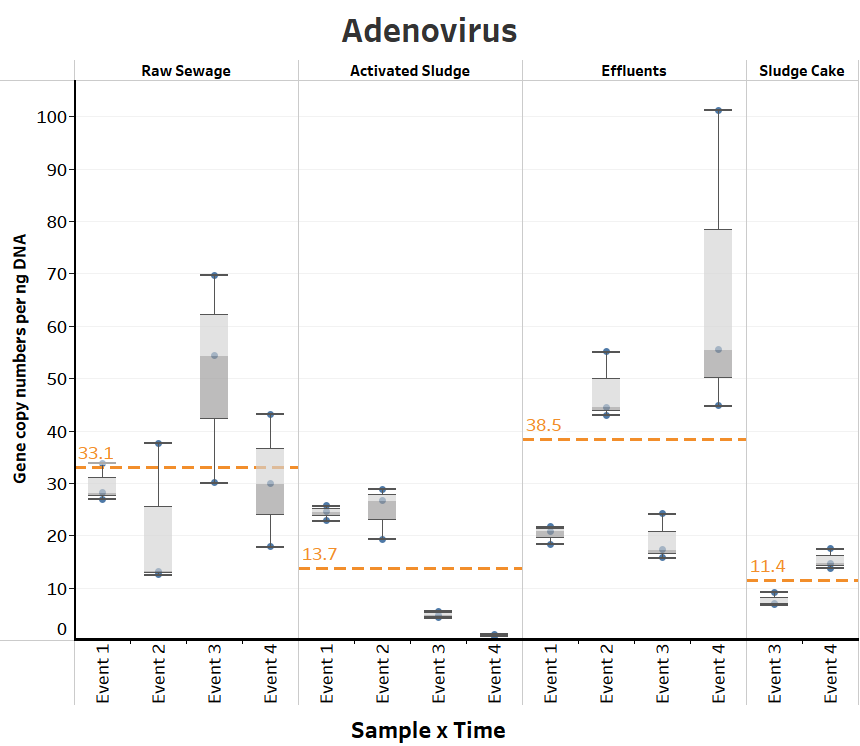
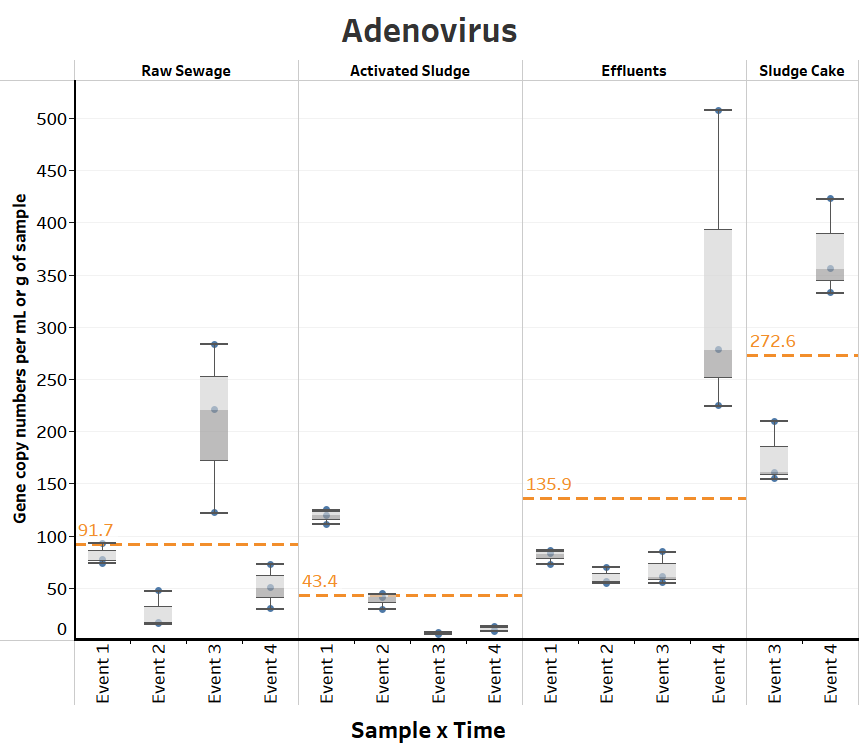






**RESULTS**

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). In some cases, these values were log10-transformed for readability purposes. 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 Figures 2-6 indicate the mean of the number of gene copies of each wastewater treatment sample across all events.



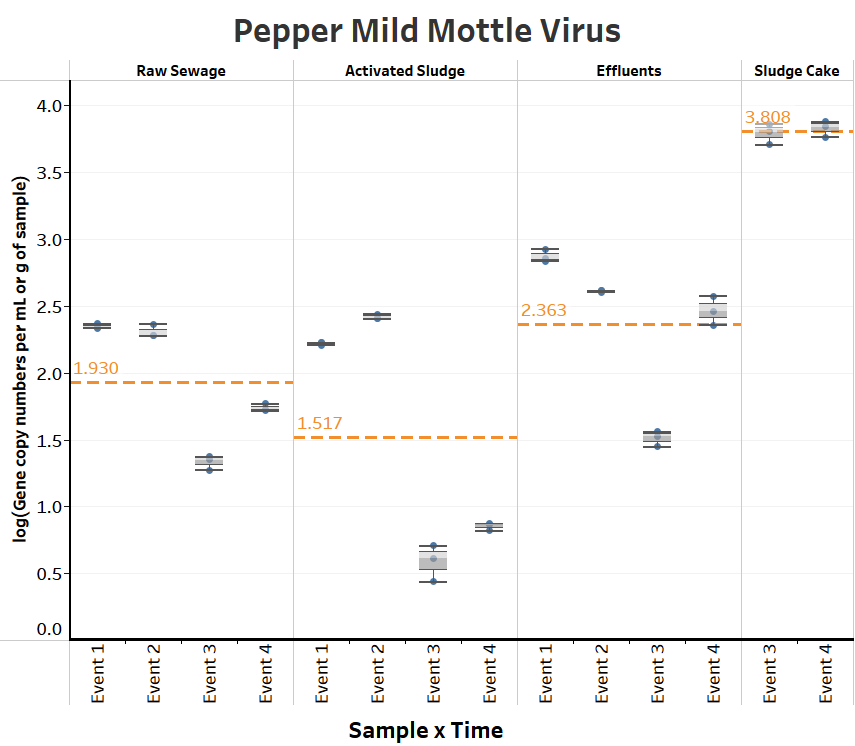
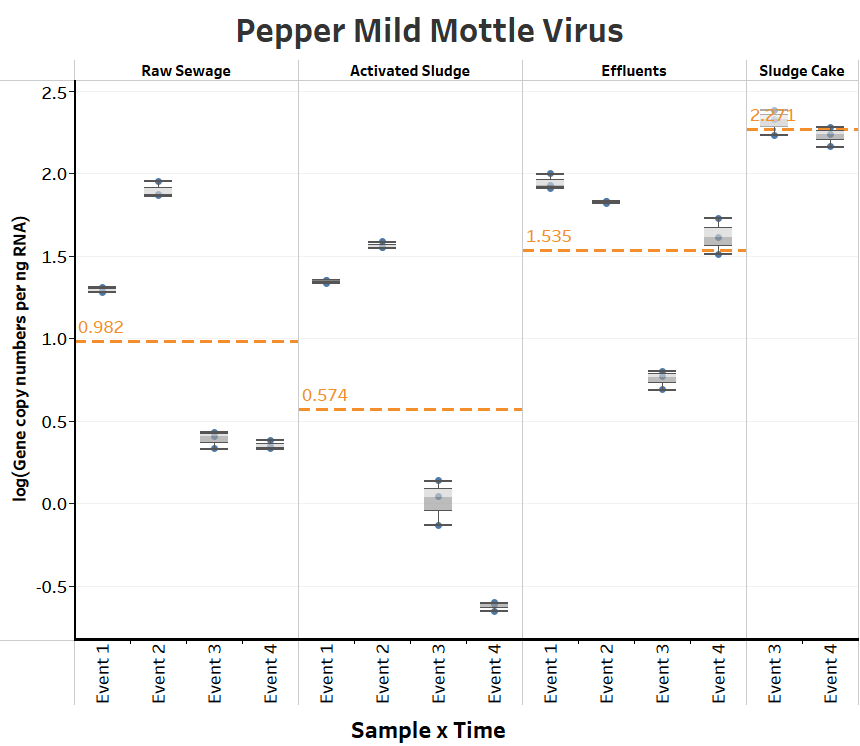
**Figure 2**. Box plots of the number of gene copies of DNA enteric viruses across each wastewater stage throughout Events 1-4. The unit for the SC in Figures 2A and 2C is gene copies per g of sample.

**2B.**

**2A.**

**2C.**

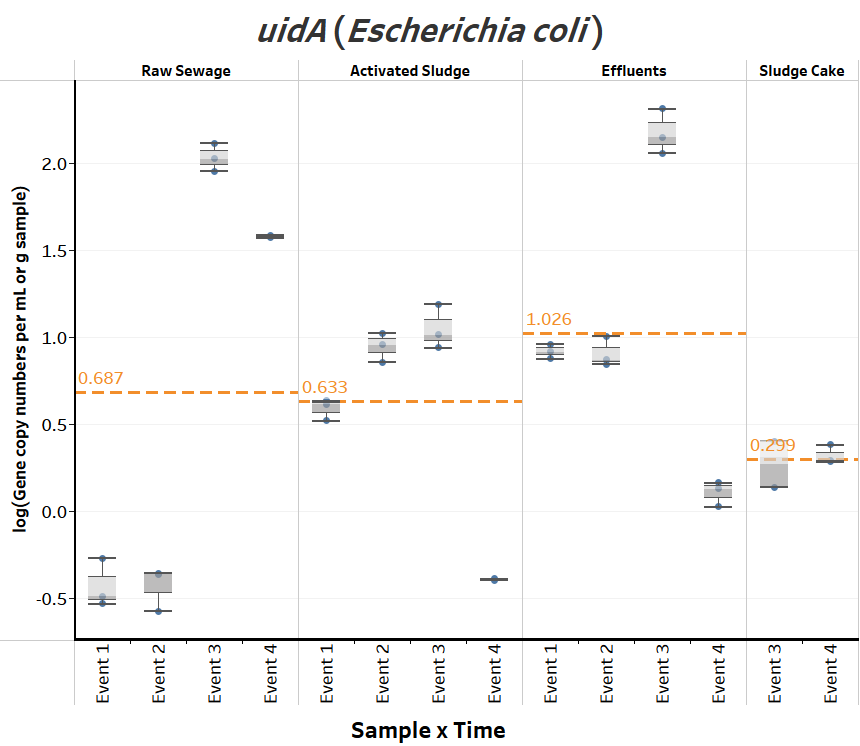
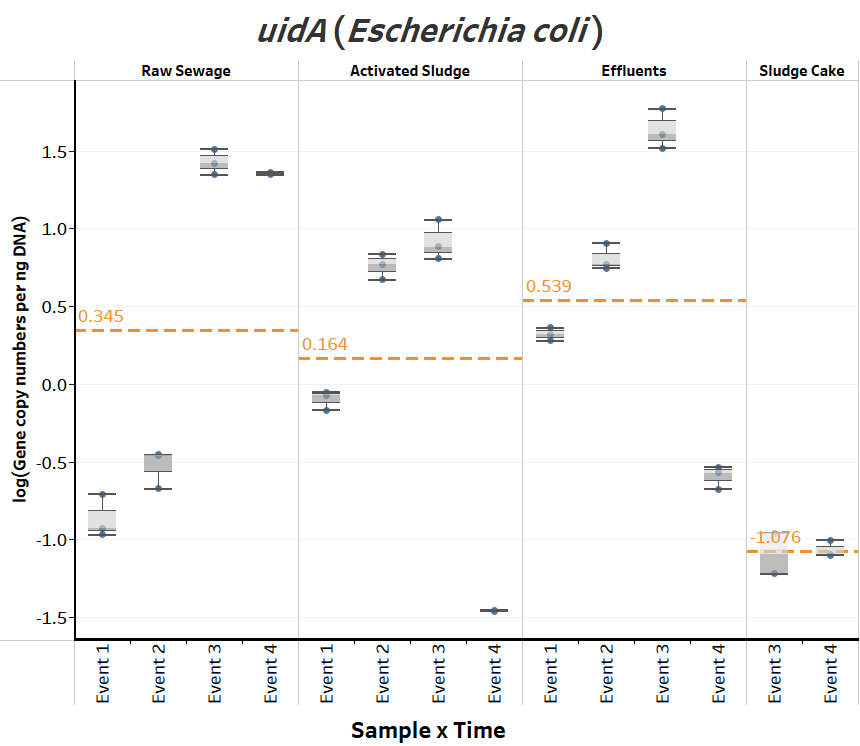
**2D.**



**3A.**

**3B.**

**Figure 3.** Box plot graphs of the number of genes copies of PMMV across each wastewater stage throughout Events 1-4. The unit for the SC in Figure 3A is gene copies per g of sample.



**4A.**

**Figure 4.** Box plots of the number of gene copies of *uidA* across each wastewater stage throughout Events 1-4. The unit for the SC in Figure 4A is gene copies per g of sample.

**4B.**

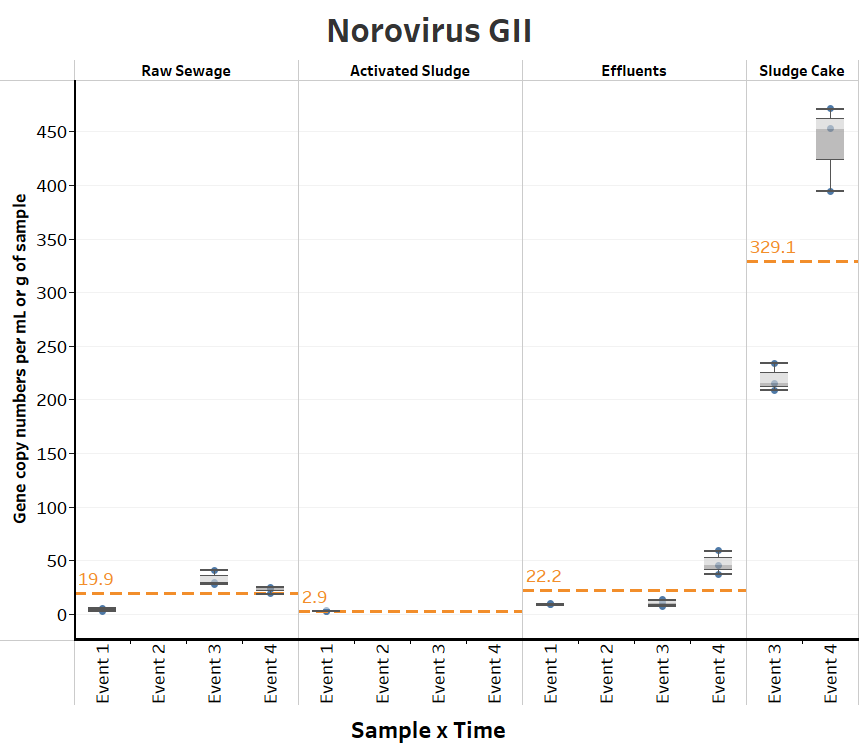
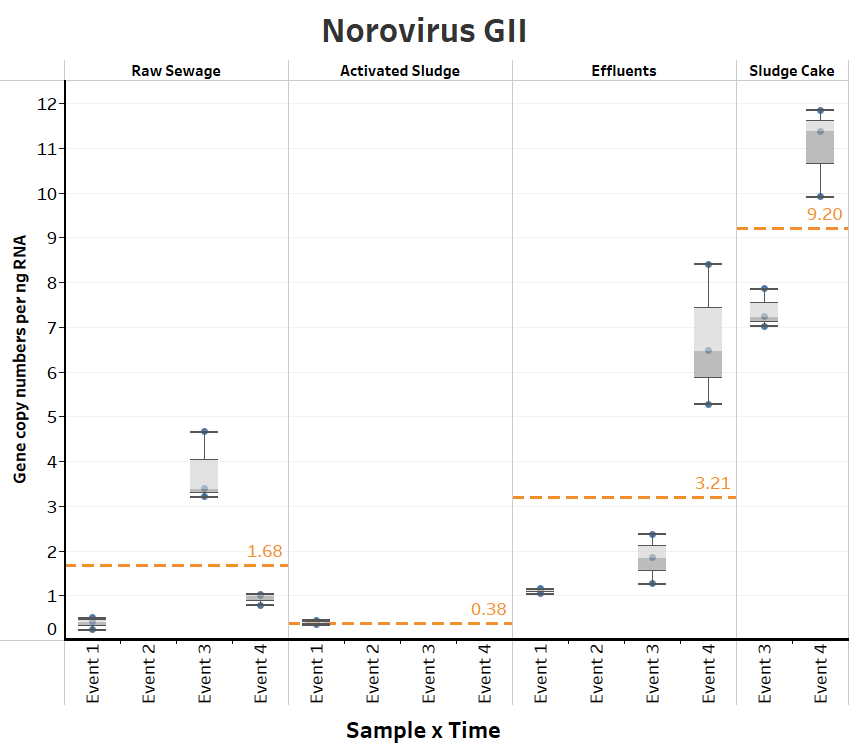
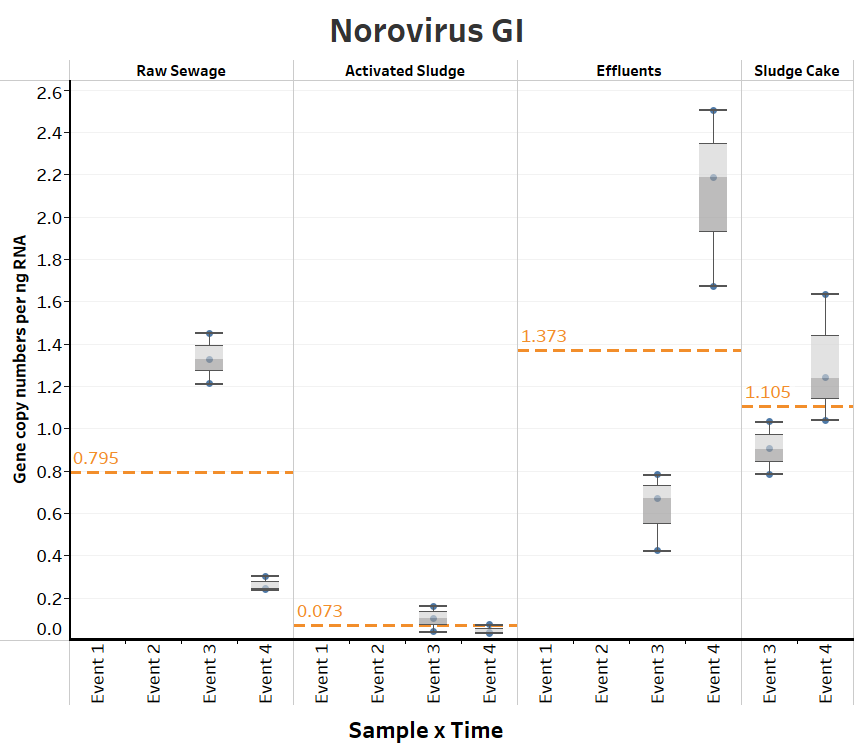
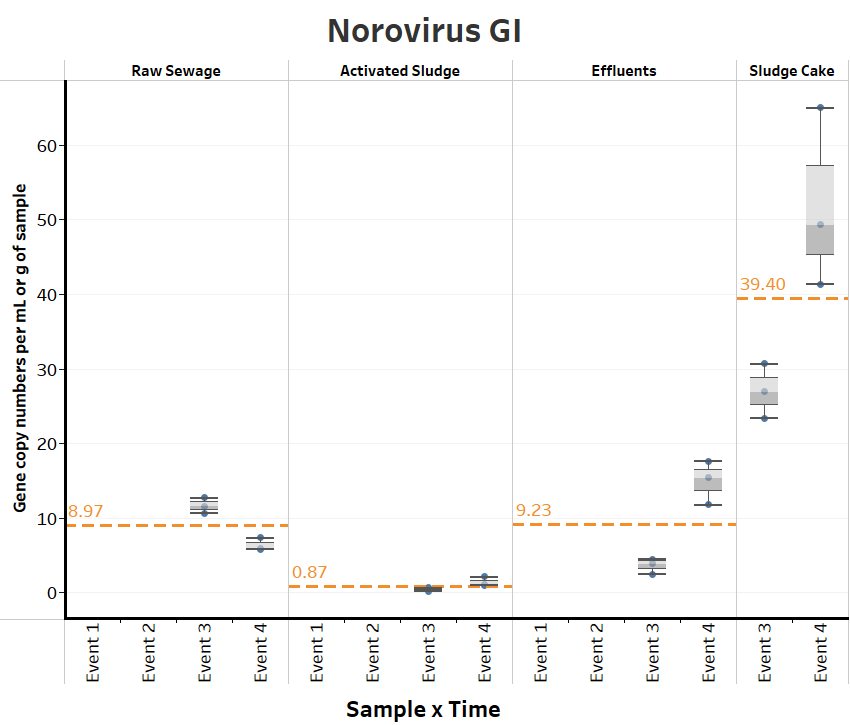
The average gene copies assessed in the RS were lower but not significantly different (p-values ranged between 0.0547 and 0.7986) than the average gene copies assessed in the EF for the DNA enteric viruses (Adenovirus and CrAssphage), PMMV, and *uidA* (*E. coli*) in terms of both volume and biomass. of 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.

[Figure 5]

**Figure 5.** Scatterplot of *uidA* GCNs in terms of sample against *E. coli* metadata measured on the same day. Each *uidA* datapoint is the average of one event’s triplicates.

To investigate potential relationships between *uidA* and *E. coli* numbers, we created a scatterplot visualizing these quantities. [Figure being revisualized in R (originally Excel) to jitter overlapping data points.] [Overall, *uidA* GCNs per mL of sample were relatively consistent to each other across events. Furthermore, there seems to be no discernible pattern between event-average *uidA* GCNs and *E. coli* MPNs for both EF and RS samples.]

[How to determine if one quantity (see *E. coli* outlier) is significantly different from the rest? Can’t use ANOVA here because there’s only one data point per event.]



**Figure S1.** Box plots of the number of gene copies of Noroviruses GI and GII across each wastewater stage throughout Events 1-4. The unit for the SC in Figures S1A and S1C is gene copies per g of sample.

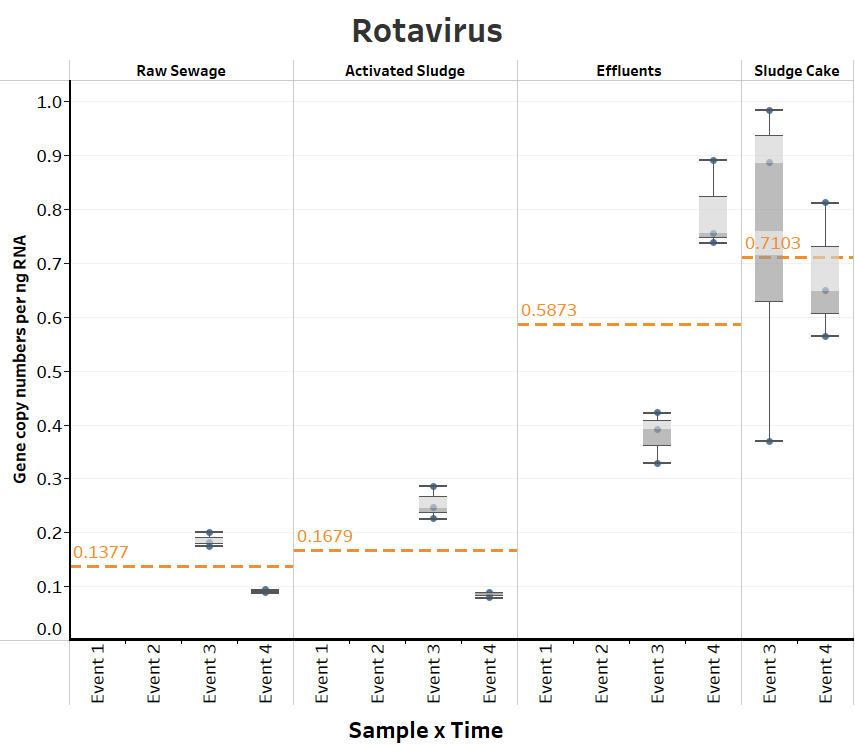
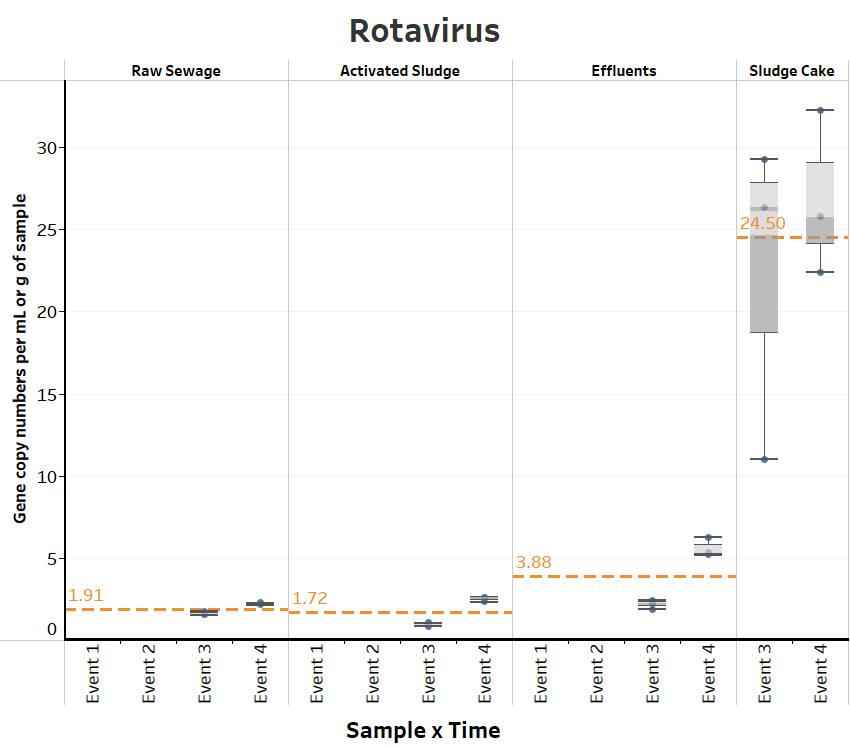
**S1A.**

**S1B.**

**S1C.**

**S1D.**

Noroviruses 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 (Figure S1). In Events 1 and 2 (fall season), gene copy quantities of Norovirus GI were below qPCR detection limits for all samples (RS, AS, and EF). Also below detection limits were Norovirus 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 Norovirus GI. There was not a significant difference between the mean gene copies of Norovirus 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).



**Figure S2.** Box plots of the number of gene copies of Rotavirus across each wastewater stage throughout Events 1-4. The unit for the SC in Figure S2A is gene copies per g of sample.

**S2A.**

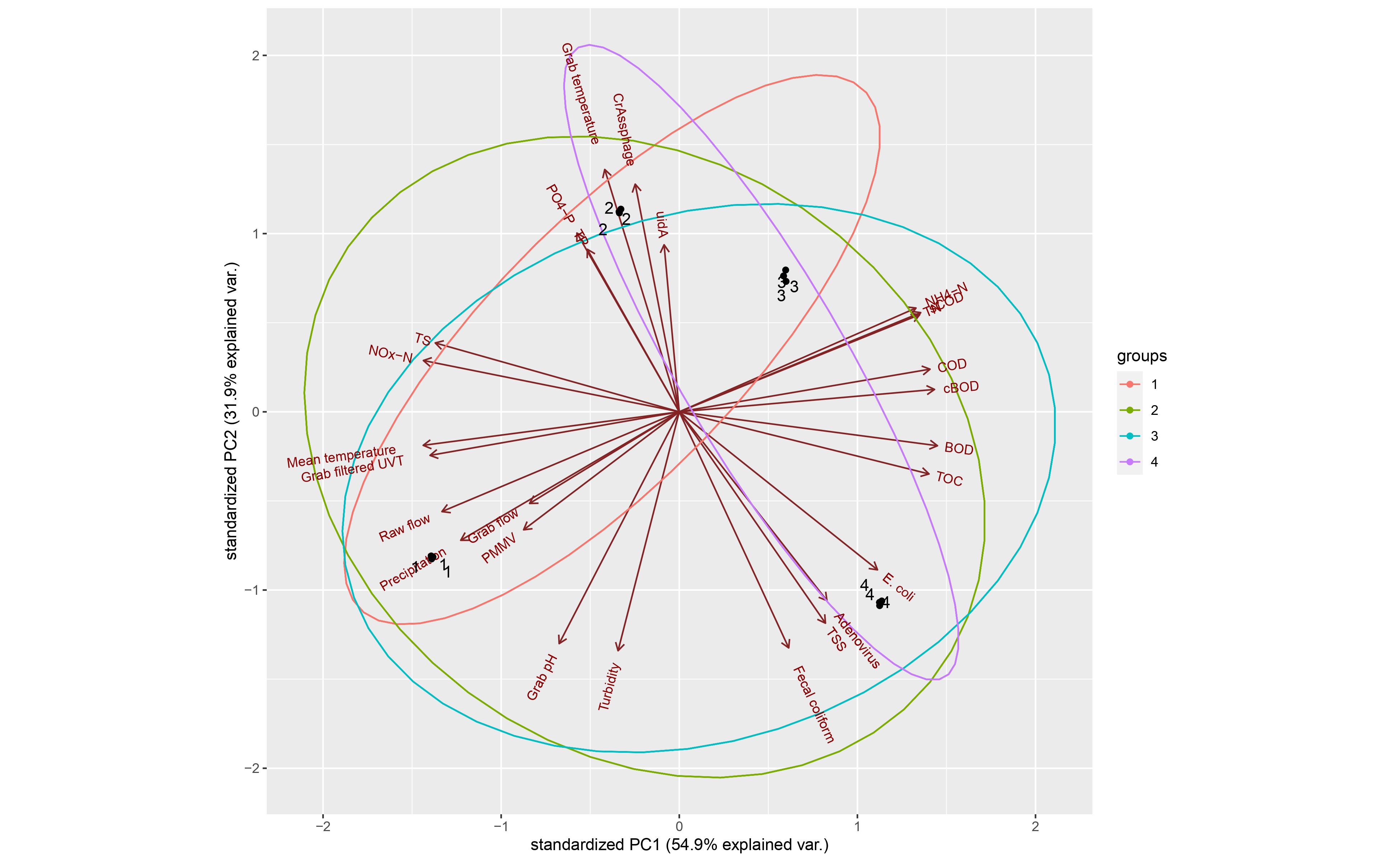
**S2B.**

In addition, we looked at Rotavirus 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 (Figure S2). Rotavirus 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.

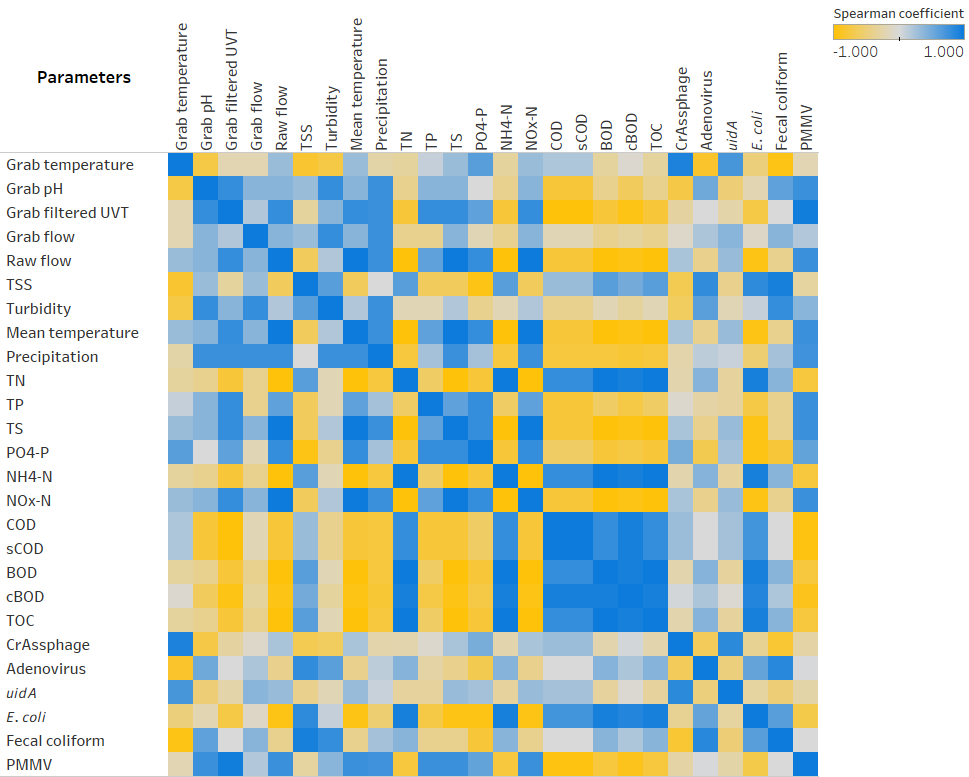
Additionally, there was no detection of gene copies for Astrovirus and Sapovirus (Sav1, Sav124, and Sav5) in any of the wastewater samples across all events.

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 number]. PC1 and PC2 were used to create the biplot in Figure 6. Biplots for PC1 versus PC3 and PC2 versus PC3 are included in the Supplementary Materials [Figure numbers].



**Figure 6.** Principal Component Analysis of log10-transformed EF parameters. Only variable not log­10-transformed was precipitation due to presence of zero values.



**Figure 7.** Heatmap showing Spearman’s rank correlation analysis between parameters collected for EF sampling events.

Overall, samples from the four events were not distinct from one another, as illustrated by the large areas of overlap between various subsets of the events. 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 (Figure 7), 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 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 Adenovirus, CrAssphage, and PMMV relative to the results for our other assays make these target more representative targets for conducting comparisons with *E. coli*. Although a reduction of Adenovirus, CrAssphage, PMMV and *uidA* GCNs was observed consistently in AS samples, 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. In a study by ElNaker, Yousef, and Hasan published in 2018, it was shown that bacterial counts in bioreactors increased as the HRT was raised from 2 to 50 hours [[Effect of hydraulic retention time on microbial community structure in wastewater treatment electro‐bioreactors - ElNaker - 2018 - MicrobiologyOpen - Wiley Online Library (oclc.org)](https://onlinelibrary-wiley-com.uml.idm.oclc.org/doi/full/10.1002/mbo3.590)]. Similarly, Pan et al. in 2004 concluded that the optimal HRT was between 2 and 12 hours, and such an HRT provided the hydraulic selection pressures favorable for the formation and maintenance of stable aerobic granules with good settleability and activity [[The effect of hydraulic retention time on the stability of aerobically grown microbial granules - Pan - 2004 - Letters in Applied Microbiology - Wiley Online Library](https://sfamjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.1472-765X.2003.01479.x)]. Therefore, it is reasonable to postulate that our target species experienced growth during AS processing. However, for each sampling event, samples were collected within a 2-hour period from RS 🡪 AS 🡪 EF consecutively. As a result, the community quantified in EF may not completely reflect the changes occurring in the AS during sample collection. Other variables to consider are the overflow of sewage from rainy events and fluctuations in mixed liquor-suspended solids (Pérez, et al. 2019).

Moreover, the duration of anaerobic sludge digestion is 15 days. In this context, GCNs of *uidA* in the SC were significantly reduced by anaerobic digestion. This may explain why the gene copies of *uidA* were lower in SC compared to all treatments. 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. This may indicate that enteric viruses are able to survive the wastewater treatment process and suggest that viral particles are being released into the aquatic environment.

Furthermore, the gene copies of CrAssphage and PMMV were highest in terms of biomass in SC. With the exceptions of CrAssphage GCN differences compared with EF (p-value = 0.30171) and RS (p-value = 0.21650), SC’s relatively greater CrAssphage and PMMV GCN was statistically significant (p-values ranged from 1.487 x 10-5 to 0.03775). Since SC is the by-product of AS and RS 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 Adenovirus in terms of biomass were not significantly different between the AS and SC samples. This reduction of viral gene copies of Adenovirus may be due to efficient digestion in the AS and the anaerobic digestion process (SC). 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 Rotavirus gene copies in the EF 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 rotaviruses in EF are found during the winter season (Li et al. 2011).

The negative results of Sapovirus (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 Sapovirus occurs during the cold and rainy months of the year. However, quantitative detection of sapovirus in wastewater and river water in Japan showed an increased concentration of sapoviruses in influents between winter and spring (December to May), but a decrease in sapovirus concentration during the summer and autumn months (July to October) (Haramoto et al. 2008). As a result, there might be other factors that can influence wastewater sapovirus concentrations.

The gene copies of GI Norovirus and GII Norovirus 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. This observation, similar to the aforementioned discussion regarding Adenovirus, CrAssphage, PMMV and *uidA*, may be related to the HRT in the NESTP facility. A possible explanation for the greatly reduced viral GCNs in AS samples is the efficiency with which viruses are removed. Many studies have shown viral removal rates for AS treatment being higher than 95% and even as high as 99.97% [[Elimination of human enteric viruses during conventional waste water treatment by activated sludge (cdnsciencepub.com)](https://cdnsciencepub.com/doi/abs/10.1139/m86-170); [Removal of bacterial and viral indicator organisms in full-scale aerobic granular sludge and conventional activated sludge systems - ScienceDirect (oclc.org)](https://www-sciencedirect-com.uml.idm.oclc.org/science/article/pii/S2589914719300763" \l "fig4); [Assessment of virus removal by a multi-stage activated sludge process - ScienceDirect (oclc.org)](https://www-sciencedirect-com.uml.idm.oclc.org/science/article/pii/0043135476900592)]. 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. The relative abundance of GI Norovirus and GII Norovirus gene copies in RS and EF during the winter months (December and February) and the absence of GII Norovirus in RS in October may be due to seasonal variability (Pérez, et al. 2019). However, the presence of GI and GII Norovirus gene copies in RS in Event 4 (February) is consistent with a study conducted by Flannery et al. (2012), in which the concentration of GI and GII Norovirus gene copies in the influents of a wastewater treatment plant were significantly higher during the winter months (January to March).

In a study conducted by Morsy El-Senousy et al. (2007), high numbers of Astrovirus 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 Astrovirus concentrations tended to decrease as temperatures increased. These results are significantly different from the quantitation of Astrovirus gene copies in the wastewater samples from the NESTP. There was zero detection of Astrovirus gene copies in all the wastewater treatment stages across all events. These results may be due to seasonal variability (Pérez, et al. 2019) and efficiency of the wastewater facility in removing this 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 not distinct from one another. This could indicate that the effects seasonal variation had on the GCNs of the various targets included in the PCA were not strong enough to clearly characterize samples from different events.

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 [[A Critical Review on Ultraviolet Disinfection Systems against COVID-19 Outbreak: Applicability, Validation, and Safety Considerations (nih.gov)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7571309/)].

**CONCLUSION**

CONCLUSION are missing here.

**ACKNOWLEDGEMENTS**

Special thanks to Dr. Qiuyan Yuan and graduate students Palwinder Singh and Kadir Yanaç in Civil Engineering at the University of Manitoba for sample collection. Research start-up funds grant number 322388 were assigned to Dr. Miguel Uyaguari-Diaz at the University of Manitoba. We also thank Dr. Brian Mark and his staff members (University of Manitoba) for the use of laboratory equipment.

**SUPPLEMENTARY MATERIALS**

**[Ideas of what to include]**

Table S1. Qubit assay results.

Table S2. All R packages used not yet mentioned.

**Figure S1.** Box plots of the number of gene copies of Noroviruses GI and GII across each wastewater stage throughout Events 1-4. The unit for the SC in Figures S1A and S1C is gene copies per g of sample.

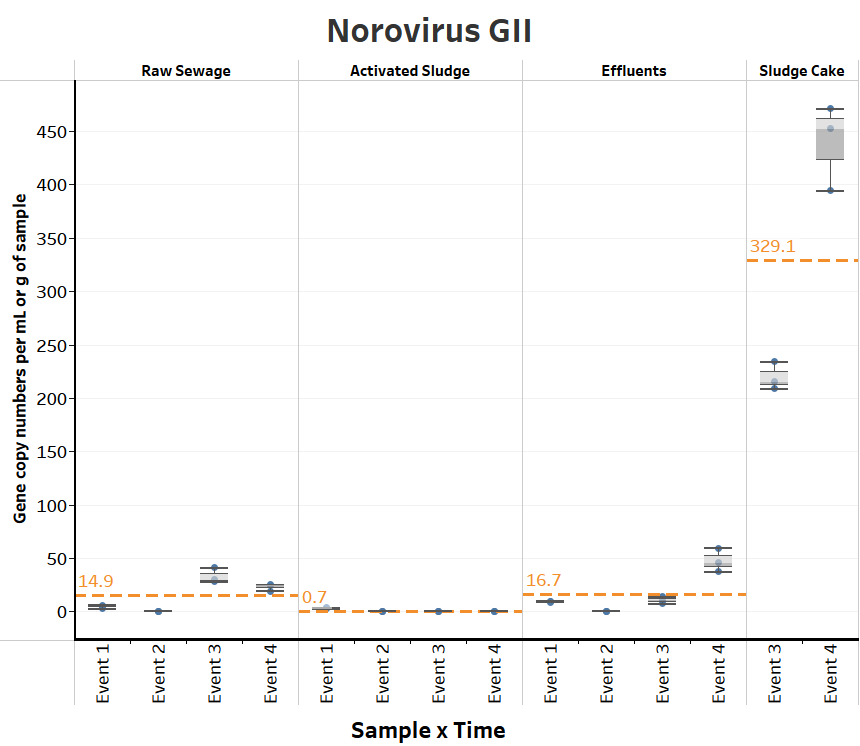
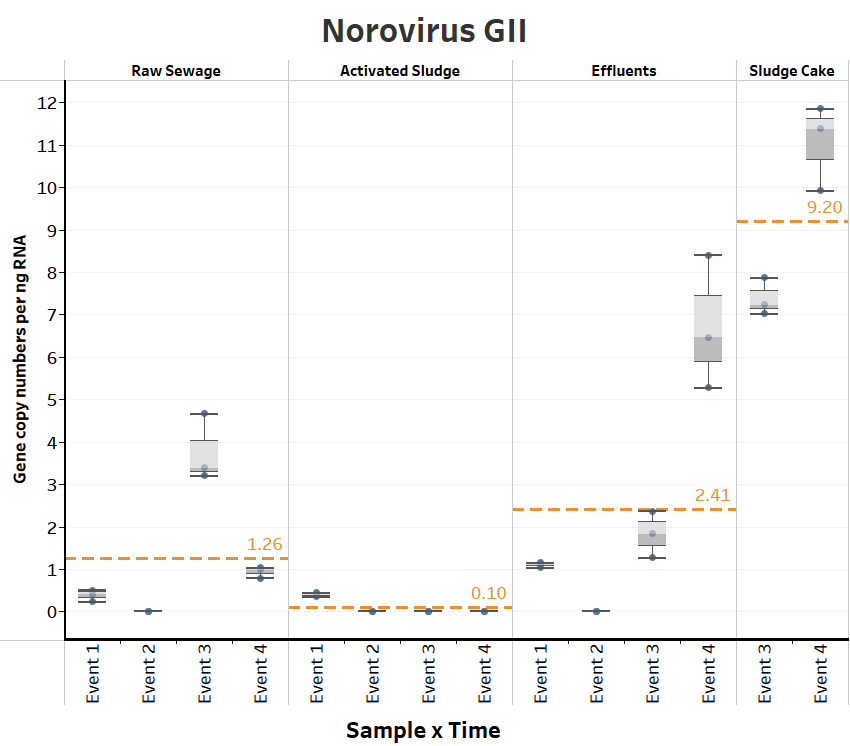
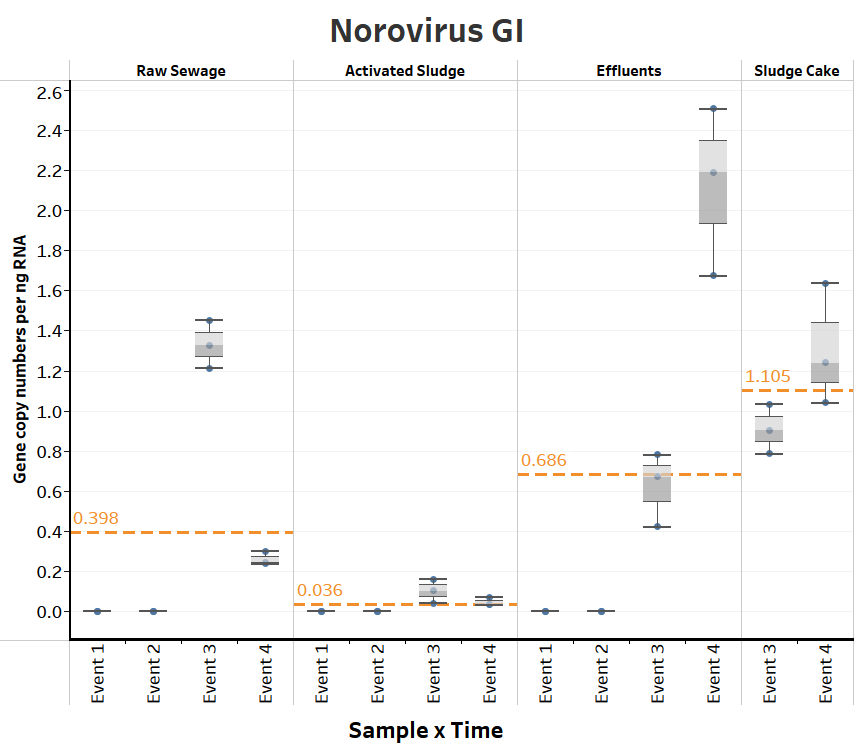
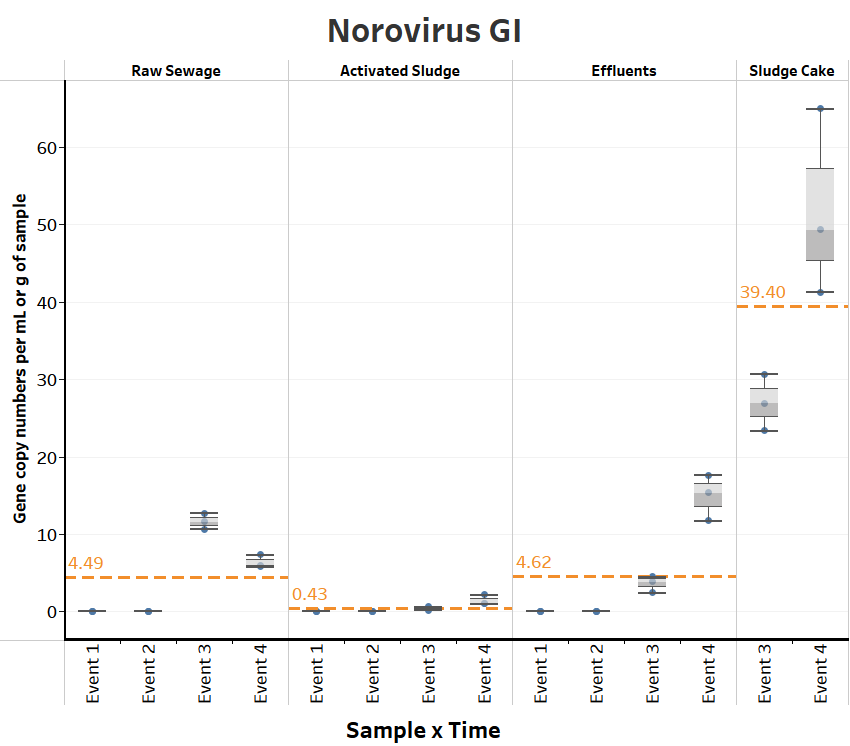
**Figure S2.** Box plots of the number of gene copies of Rotavirus across each wastewater stage throughout Events 1-4. The unit for the SC in Figure S2A is gene copies per g of sample.

Table S4. Summary of weight of components.

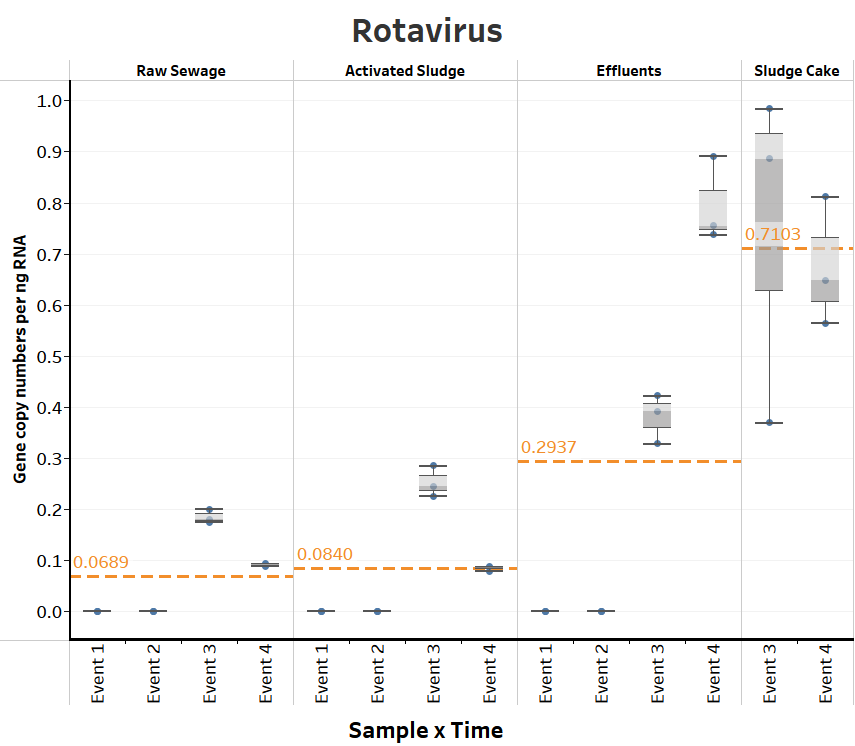
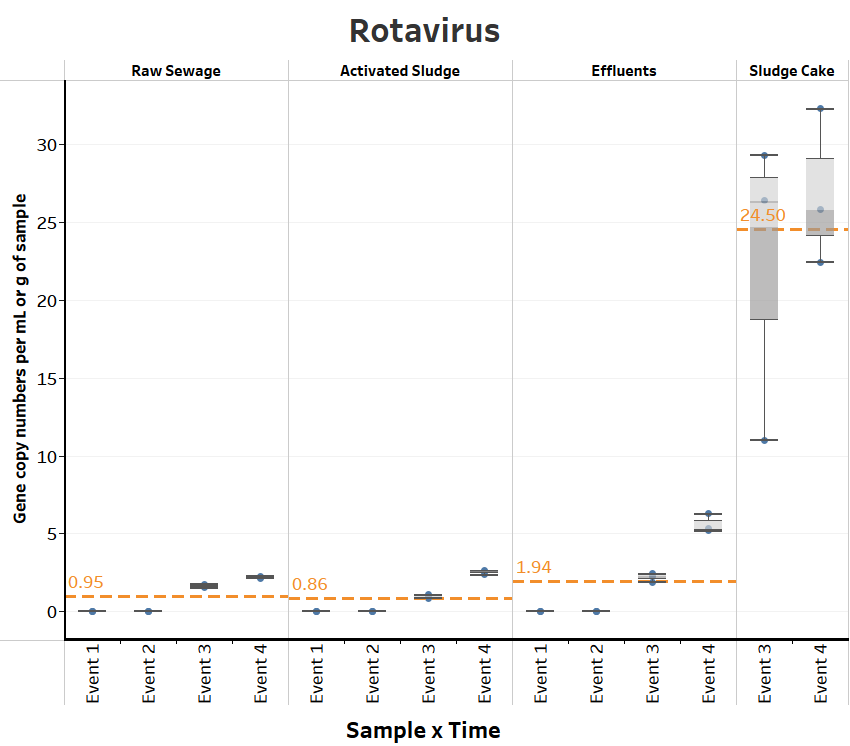
Figure S. PC1 vs PC3 biplot.

Figure S. PC2 vs PC3 biplot.

**BACK-UP OF OLD FIGURE VERSIONS**



**Figure S1 - OLD.** Box plots of the number of gene copies of Noroviruses GI and GII across each wastewater stage throughout Events 1-4. The unit for the SC in Figures S1A and S1C is gene copies per g of sample.



**Figure S2 - OLD.** Box plots of the number of gene copies of Rotavirus across each wastewater stage throughout Events 1-4. The unit for the SC in Figure S2A is gene copies per g of sample.

**S2A.**

**REFERENCES**

1. Atabakhsh, P., Kargar, M., and Doosti, A. 2020. Molecular detection and genotyping of group A rotavirus in two wastewater treatment plants, Iran. *Brazilian Journal of Microbiology*. 51(1):197–203. doi: 10.1007/s42770-019-00131-0.
2. City of Winnipeg. 2019a. Sewage treatment plants: Introduction [online]. Available from https://www.winnipeg.ca/waterandwaste/sewage/treatmentPlant/default.stm#tab-introduction [accessed 6 April 2020].
3. City of Winnipeg. 2019b. Sewage treatment plants: North End Sewage Treatment Plant (NEWPCC) [online]. Available from https://www.winnipeg.ca/waterandwaste/sewage/treatmentPlant/default.stm#tab-north-end-sewage-treatment-plant [accessed 6 April 2020].
4. Dutilh, B.E., Cassman, N., McNair, K., Sanchez, S.E., Silva, G.G., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., et al. 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* 5(4498): 1−11. doi: 10.1038/ncomms5498.
5. Flannery, J., Keaveney, S., Rajko-Nenow, P., O’Flaherty, V., and Dore, W. 2012. Concentration of norovirus during wastewater treatment and its impact on oyster contamination. doi: 10.1128/aem.07569-11.
6. Government of Canada. 2019. Historical Data [online]. Available from <https://climate.weather.gc.ca/historical_data/search_historic_data_e.html> [accessed 6 April 2020].
7. Haramoto, E., Katayama, H., Phanuwan, C., and Ohgaki, S. 2008. Quantitative detection of sapoviruses in wastewater and river water in Japan. Letters in Applied Microbiology, 46(3): 408-413. doi: 10.1111/j.1472-765X.2008.02330.x.
8. Jumat, M.R., Hasan, N.A., Subramanian, P., Heberling, C., Colwell, R.R., and Hong, P. 2017. Membrane Biorector-Based Wastewater Treatment Plant in Saudi Arabia: Reduction of Viral Diversity, Load, and Infectious Capacity. *Water* 9(7): 534. doi:10.3390/w9070534.
9. Li, D., Gu, A., Zeng, S.Y., Yang, W., He, M., and Shi, H.C. (2011) Monitoring and evaluation of infectious rotaviruses in various wastewater effluents and receiving waters revealed correlation and seasonal pattern of occurrences. J Appl Microbiol 110(5): 1129–1137. doi: [10.1111/j.1365-2672.2011.04954.x](https://doi.org/10.1111/j.1365-2672.2011.04954.x).
10. Morsy El-Senousy, W., Guix, S., Abid, I., Pinto, R., and Bosch, A. 2007. Removal of Astrovirus from Water and Sewage Treatment Plants, Evaluated by a Competitive Reverse Transcription-PCR. *Applied and Environmental Microbiology*, 73(1): 164–167. doi: 10.1128/AEM.01748-06.
11. Pérez, M.V., Guerrero, L.D., Orellana, E., Figuerola, E.L., and Erijman, L. 2019. Time Series Genome-Centric Analysis Unveils Bacterial Response to Operational Disturbance in Activated Sludge. *MSystems*, 4(4): e00169-19. doi: 10.1128/mSystems.00169-19.
12. Ritalahti, K.M., Amos, B.K., Sung, Y., Wu, Q., Koenigsberg, S.S., and Löffler, F.E. 2006. Quantitative PCR Targeting 16S rRNA and Reductive Dehalogenase Genes Simultaneously Monitors Multiple *Dehalococcoides* Strains. *Applied and Environmental Microbiology,* 72(4): 2768. doi: 10.1128/AEM.72.4.2765-2774.2006.
13. Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., and Shanks, O.C. 2017. Quantitative CrAssphage PCR assays for human fecal pollution measurement. *Environ. Sci. Technol*. 51(16): 9146-9154. doi: 10.1021/acs.est.7b02703.
14. Varela, M., Ouardani, I., Kato, T., Kadoya, S., Aouni, M., Sano, D., and Romalde, J.L. 2018. Sapovirus in Wastewater Treatment Plants in Tunisia: Prevalence, Removal, and Genetic Characterization. *Applied and Environmental Microbiology*. 84(6): e02093–17–17.atom. doi: 10.1128/AEM.02093-17.
15. Zhang, T., Breitbart, M., Lee, W., Run, J., Wei, C., Soh, S., et al. (2006). RNA Viral Community in Human Feces: Prevalence of Plant Pathogenic Viruses. *PLoS Biology*. 4(1): e3. doi: 10.1371/journal.pbio.0040003.