Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

Pol. J. Environ. Stud. Vol. 30, No. 4 (2021), 1-13
DOI: *10.15244/pjoes/127888* ONLINE PUBLICATION DATE:

Original Research

Viral Communities Distribution and Diversity in a Wastewater Treatment Plants Using High-Throughput Sequencing Analysis

Cecilia Oluseyi Osunmakinde[1], Ramganesh Selvarajan[2]*, Bhekie Mamba[1], Titus A.M Msagati[1]**

[1]Nanotechnology and Water Sustainability Research Unit, College of Science, Engineering and Technology, University of South Africa-Science Campus, Florida 1710, South Africa
[2]Department of Environmental Sciences, College of Agriculture and Environmental Sciences, University of South Africa-Science Campus, Florida 1710, South Africa

Received: 12 June 2020
Accepted: 26 September 2020

Abstract

Waterborne pathogens are still considered a threat to human health. Wastewater treatment plants (WWTPs), contain a high populace and diversity of viruses that can significantly affect the aquatic ecosystem. Subsequently, the occurrence and recurrence of viral pathogens are alarming that exist in human populations, which are potentially discharged into sewage systems. This study investigated the distribution and diversity of the viral communities from three WWTPs in Gauteng Province, South Africa using next-generation sequencing technology. The results uncovered that most of the viral populaces in the wastewater plants belong to the families *Siphoviridae*, *Microviridae,* and *Myoviridae* of the order Caudovirales. Other families including *Baculoviridae*, *Partitiviridae, Virgaviridae*, and *Tymoviridae* were likewise recorded. On the other hand, enteric pathogens detected were *Astroviridae, Reoviridae,* and *Coronaviridae.* The predicted functional features for the viromes revealed that the DNA metabolism was the major function within the community followed by carbohydrate and amino acid metabolism, suggesting that the treatment plants provide a rich biological environment for the growth of diverse viral species and other processes. The overall results in this study highlight the biological significance of WWTPs as a hotspot for some viral agents. The contamination routes of these viral agents are needed for efficient and cost-effective management strategies in water systems. Further, viral metagenomics ought to be considered as a key approach for revealing the diversity and early warning of emerging disease outbreaks.

Keywords: metagenomics, pathogens, public health, viral ecology, wastewater treatment plant

*e-mail: ramganesh.presidency@gmail.com
**e-mail: msagatam@unisa.ac.za

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

Introduction

Clean water is essential for the nourishing lifespan of all aquatic and terrestrial organisms, including humans. It is an important resource for the survival of humans as well as other living organisms and deserves proper monitoring to protect it from various sources of contamination [1]. Wastewater treatment plants (WWTPs) are enormous bio-processors that acquire wastes from households, manufacturing, agriculture, environmental run-off and other drain channels, which make them reservoirs for numerous pollutants and pathogenic microbes including bacteria, protozoan and viruses [2-6]. Generally, over the years, it has become progressively noticeable that viruses are a major health concern within the microbial populations as they are the leading cause of most waterborne infections [7]. Their presence has been demonstrated in both treated and untreated wastewater, WWTPs do not only pull together pathogenic viruses from various sources in the environment, but they also tend to discharge a number of viruses into other aquatic ecosystems [4]. The viral communities in WWTPs have also revealed large numbers of bacteriophages that are specific to bacterial hosts are thought to strongly influence bacterial community structure, dynamics and threat to human health by facilitating the buildup of antibiotic resistant bacteria and antibiotic resistance genes [8].

Viruses are often regarded as the most subtle and dangerous of all waterborne pathogens found in wastewater systems [2, 8, 9], with the most commonly detected human viral pathogens including human Adenoviruses, Norovirus, Astrovirus, Sapovirus, Enteroviruses, Rotaviruses, Astroviruses, Reoviruses, Hepatitis A, and Caliciviruses [10-15]. These enteric viruses are usually shed from faeces, urine, and respiratory secretions from infected host and then released into the sewage system. When humans or animals drink or get in contact with the water from different sources it could give rise to different adverse health implications and sometimes huge economic loss [16]. When pathogens are not adequately removed during the treatment process at the WWTPs, they are then consistently released into natural watersheds and drinking supplies where many of them can persist for a longer period of time [15]. Although viruses in general are found in low concentrations in the environment, the concentration levels in different WWTPs effluents may vary from one plant to another depending on the type of treatment, geographical location and the diverse activities been carried out within the surroundings [16]. However, up to date, no regulations have been implanted to monitor viral concentrations in wastewater before it is discharged into the environment [1].

In recent years, enteric viruses have been fully studied, monitored and detected in different parts of the world in both environmental and clinical samples. Although they are generally more difficult to detect in environmental samples [2, 8], a number of enrichment techniques have been established in the literature that overcomes this perception [17–19]. As harmful as viruses are, they are important to guarantee satisfactory water quality as the traditional bacterial indicators are not always appropriate for the prediction, determination and occurrence of viral pathogens at treatment plants [20]. Though some bacteriophages have been recommended as indicators of viral contamination, their presence does not always show any significant relationship with the occurrence of enteric viruses [20-23].

A number of studies have considered cell culture as a gold standard for the detection of viral species/strain [10, 24-26]. However, it has been recognized that cell culture does not allow certain for the replication of all viral species and strains, as a number of mechanisms and conditions control the efficiency of the technique [26, 27]. Molecular approaches based on nucleic acid amplification and other techniques have also been considered and demonstrated as a sensitive and better approach for the detection of viral species, but this has its own shortcomings such as the use specific reference sequences for primer design [11, 18, 19, 25, 29]. The launch of next generation sequencing (NGS) has been instrumental in the strategies for the identification of known, new viral genomes and pathogens through viral metagenomics from different environmental matrices [19, 30-36]. Such high throughput sequencing techniques have proven to be a suitable approach for analyzing environmental viromes because most viruses lack a universal gene marker which can target specific regions [36]. This allows for the simultaneous detection of multiple viruses including both DNA/RNA viruses without preference to any particular gene; the overall microbial domain composition, functional analysis and possible pathogen identification are also described. Over the years most of the studies in South Africa have focused mainly on the detection and monitoring of specific human pathogenic viruses and bacteriophages in wastewater and other environmental samples using molecular techniques, many of these studies do not include any diversity, functional analysis, and other pathogenic viral studies [1, 10, 37, 38]. In the study, we characterize the presence of viromes within the influent and effluent of three wastewater treatment plants in Gauteng, South Africa using high throughput sequencing approach. The results obtained provided an extensive view of the different viral families, genetic diversity, functional distributions, plant, animal and human pathogenic viruses present at the WWTPs which might be of great concern to the ecosystem. The independent identification of viral pathogen diversity in WWTP would represent a significant advance in water quality regulation and treatment design decisions for water management teams. Also, it could be used to monitor specific viruses, which may complement the bacterial indicators that are already in use.

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

Materials and Methods

Sample Collection

Three WWTPs in the province of Gauteng, South Africa were selected for this study as it is one of the provinces with high population and urbanization rate. Influents and effluents samples were collected from the WWTPs (IDP, EDP, IFP, EFP, IPP, and EPP). Plant A is located in Pretoria, a drinking water treatment plant in Tshwane municipality that serves approximately 2.125 million residents of Pretoria west, treating both raw sewage and industrial wastes. Wastewater from the central Pretoria area is collected in a main outfall sewer that runs through the Daspoort wastewater treatment works with a capacity of about 60 Mℓ/day. The treatment process consists of a comprises of a screening and grit removal channels, primary sedimentation tanks (Dortmund tanks), biological filtration units, sludge digestion in rectangular tanks while disinfection is done with chlorine prior to discharge of the treated effluent into the Apies River that flows down to the Pienaars River. Plant B is located in Krugersdorp Johannesburg and treats wastes from approximately 362, 422 of the people; the plant receives its wastes from households, hospital and industries. The treated effluents from the plant are released into the Bloubankspruit River which is connected to the Crocodile River which then flows into the Hartbeespoort Dam. The dam is a major source of tourist attraction in the Northwest province of South Africa. Plant C is also located in Johannesburg around the Mogale city municipality which receives mainly domestic sewage from households and industrial wastes with an average population of 140, 643 residents. After treatment, the effluents are released into the upper Wonderfonteinspruit, which is part of the Mooi River catchment system, which flows into the Vaal Dam. It is a major dam that supplies drinking water to most people in Johannesburg and its surrounding areas through Rand water. Plant B and C have a capacity of about 32 Mℓ/day, with a similar treatment process consisting of a primary treatment, a secondary treatment (activated sludge), and disinfection with chlorine like Plant A.

Approximately about 2,000L of water was pumped into 2000L JoJo tanks that had already been sterilized with 0.525% sodium hypochlorite. The collected wastewater samples were immediately transported to the University of South Africa (UNISA, Florida Campus, RSA) College of Agriculture and Environmental studies (CAES) laboratory for further processing. All field meters and equipment were checked and calibrated according to the manufactures specification period. The physicochemical parameters such as temperature, pH, conductivity, salinity, total dissolved solids (TDS), and dissolved oxygen (DO), ammonia nitrogen (NH_3-N), were then measured and recorded on-site using portable multi-parameter meter YSI professional plus (Xylem Inc., USA). The obtained data were then subjected to descriptive statistical analysis at 95 % confident limit.

Concentration of Viruses in Wastewater Samples

A total of six wastewater samples (influent and effluent samples each from the three plants) were collected and concentrated using an electropositive filter method as described previously [39]. Briefly, wastewater samples were primarily concentrated using a 10-inch NanoCeram^R 0.2 micron pleated electropositive filter (Argonide Corporation Sanford, FL USA) at a flow rate 2.5 L/min. The filter was housed in a cylinder module and water passed through until the membrane filter fouled beyond the point at which water could no longer flow through the filter using a previously described method [9, 39] with slight modification. Elution of the electropositive filter was achieved by adding about 1 L of a mixture 1.5% beef extract and 0.05 M glycine pH 9.5, while a vacuum pump was used to force the eluent out and then the eluent was subjected to acid precipitation with 1 M HCl and the pH adjusted to 3.5. The resulting solution was centrifuged at 2,500 × g at 4ºC for 15 min, the pellet formed was completely dissolved using 30 mL 0.15 M sodium phosphate buffer solution (Na_2HPO_4• $7H_2O$, pH 9.5). The solution was mixed until the pellets dissolved completely and the pH of the solution adjusted with a 1M NaOH. The solution was then placed into a 50 mL centrifuge tube and centrifuged at 7,500 x g for 10min at 4ºC. The supernatant was collected using a pipette and then poured into a new 50 mL centrifuge tube while the pH is adjusted to 7.2 with a 1M HCL for the stabilization of the viral particles and the pellet discarded. The resulting supernatant was then loaded into a 60 mL syringe and passed through a 0.22µm pore size syringe filter to remove non-viral organisms and finally stored at -80ºC. The viral particles from the primary concentrate were further concentrated to approximately 1 mL by centrifugation through the use of Centripep Ultracel YM-30 kDa cutoff centrifugal filter units according to the manufacturers protocol (Millipore Sigma™ USA), and stored at -80ºC until further analysis.

Extraction of Viral Nucleic acid and Reverse Transcription (RT)

The viral DNA and RNA were extracted from the concentrated wastewater samples using the Qiagen AllPrep Power Viral DNA/RNA kit (Qiagen, Hilden, Germany) to obtain a final volume of 70 µL, according to the manufacturer's protocol. Following extraction, the purity and yield of extracted genomic of viral DNA/RNA extracts from all samples were verified for quality control purposes using the Biodrop µLite spectrophotometer (Biochrom, USA), concentrations were measured using the Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA USA). The RT reaction was performed using the Qiagen QuantiNova™ Reverse Transcription kit (Qiagen,

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

Hilden, Germany). Briefly, purified RNA sample was incubated in QuantiNova gDNA Removal Mix at 45ºC for 2 minutes to effectively reduce contaminating genomic DNA. In the reverse-transcription step, to 10 µL of the extracted RNA were added to 1 µL, of reverse transcription enzyme, 4 µl reverse transcription mix, 2 µl gDNA removal mix, RNA, 1 µl RNase-Free water was thawed briefly and then incubated for 2 min at 45ºC, and then place immediately on ice according to the manufacturer's instructions. The reaction mixture (20 µL) was briefly vortexed to ensure total mixing and then centrifuged. The samples were placed into PCR tubes in the cycler repeatedly to continue the annealing for 3 min 25ºC, reverse-transcription step 10 min 45ºC, inactivation of reaction for 5 min 85ºC. The resulting 20 µL of cDNA was stored at -20ºC until further experimental procedures.

Illumina Next Generation Sequencing

The cDNA extracts were fragmented and sequenced using the Illumina HiSeq 2500 technologies (Illumina HiSeq, Roche Technologies) at the Agricultural research council (ARC) Biotechnology platform, South Africa. Whole genome sequencing library preparation was performed using the Illumina TruSeq stranded total RNA sample preparation kit following manufacturer's protocols (Illumina, San Diego, CA, USA). Briefly 50ng of genomic DNA was first tagmented in a transpose-mediated reaction that simultaneously fragments and tags DNA with adapters. The adapter tagged DNA fragment libraries were purified to remove any unwanted contaminants from the tagmentation reaction. After preparation, libraries experienced quality control and were quantified using Qubit dsDNA fluorometer assays and all samples passed quality control. The samples were pooled together and the pool was loaded on one lane of an Illumina HiSeq 2500 system flow cell using v4 chemistry (2 x 125bp) paired end format using Illumina Rapid SBS reagents, the sequence output from the Illumina was demultiplexed and converted to FASTQ format with Illumina scripts.

Bioinformatics and Data Analysis

The general bioinformatics approach was firstly, to trim and clean sequencing reads to remove the lower quality bases of the paired-end reads. The base calling was done on the Illumina Real Time Analysis (RTA) system, and then converted into a FASTQ format. Quality control was performed on the raw sequences using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) program. The obtained data were processed with Trimmomatic v. 0.36 online software for the removal of adapters, barcodes, primer mismatches in the sequences and also to filter out all low-quality reads according to [40]. After trimming, the reads were assembled into contig to decrease the possibilities of false positive detection, which was achieved by using an iterative de Bruijn graph de novo assembler for short reads by sequencing data with specified parameters [41]. After the assembling the sequences into contigs, the viromes were uploaded unto METAVIR web server for viral taxonomic composition, genetic richness and diversity of the viruses at the treatment plant [42]. The viral taxonomic composition in each sample was computed from a BLAST comparison with the Refseq complete viral genomes protein sequences database from NCBI using tBLASTx with a maximum e-value of 10^{-3} whereby each protein is associated to its best hit from BLAST [29, 30]. Predicted proteins from contigs were then compared to Refseq through BLAST, and then each protein is affiliated to its best BLAST best hit using GAAS to estimate the number of viral particles for each viral species in the sample [43]. Each sequence was classified into its likely taxonomic group; whereby the species nomenclature and classification were performed according to the NCBI Taxonomy database standards based on the best BLAST result at a 90% coverage cut-off. Taxonomy was assigned by Kraken [44], and only the reads that could be mapped to specific genus with high confidence were considered for genome mapping. For the visualization of the taxonomy, heatmaps were generated using ggplot2 in R platform. Potential human-associated pathogens were determined based upon their phylogenetic relationship with previously discovered pathogens and complete viral genomes from the GenBank sequence database at NCBI. Predicted proteins from contigs are compared to Refseq through BLAST, and then each protein is affiliated to its best BLAST hit. The MG-RAST server produces automated taxonomic assignments using BLASTx searches against the SEED non-redundant database and other accessory databases to produce some metabolic profiles of the viral metagenomes by BLASTX comparisons using the SEED-Subsystem data set.

Results and Discussion

Physicochemical Analysis

The results of the selection of few physicochemical parameters of samples (IDP, EDP, IFP, EFP, IPP, and EPP) from the different points are as shown in Table 1. The pH levels of the samples varied significantly in the sample points throughout the study period and ranged from 6.64 to 7.8 in the influent; 7.05 to 8.12 in the effluent samples, with an average temperature of 15.65ºC. Additional physicochemical profiling parameters such as dissolved oxygen, electrical conductivity, dissolved organic carbon, and salinity conspicuously varied among the wastewater samples. More variation of the conductivity was observed in the water samples, with the values ranging from 685.5-902 μScm^{-1} in influent, while effluent samples were 506-1016 μScm^{-1}. The dissolved organic carbon, dissolved oxygen and salinity values varied among the

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

Table 1. Selected physicochemical parameters of influent and effluent water samples at three wastewater treatment plants (WWTPs). (Mean±SD, n = 3).

Parameters	Influent				Effluent			
	IDP	IFP	IPP	*P-value*	EDP	EFP	EPP	*P-value*
pH	6.64±0.01	7.8±0.01	7.2±0.01	0.0028	7.05±0.03	8.12±0.06	7.2±0.01	0.144
Conductivity (μScm^{-1})	840.5±2.1	902±5.6	685.5±6.4	<0.0001	506±0.01	774±2.8	1016±7	0.0085
Salinity (µg/L)	0.42±0.01	0.45±0.01	0.34±0.01	<0.0001	0.25±0.01	0.38±0.01	0.51±0.01	<0.0001
Temp ºC	21.2	15.9	18.4	<0.0001	19.9	15.6	18.5	0.0002
TDS (mg/L)	420.5±0.7	451.5±4	513.3±18	0.0003	253.0 ±0.1	387.5 ±2	508.0 ±4.2	<0.0001
DO (mg/L)	0.44±0.6	0.67±0.1	0.48±0.4	<0.0001	3.45±0.4	3.04±0.5	1.89±0.8	0.0001
DOC (mg/L)	20.39±0.62	55.13±0.8	154.93±2.5	0.0001	5.59±0.7	11.73±0.7	18.16±0.9	0.033
Phosphate (mg/L)	5.07±0.6	12.72±0.1	23.29±0.1	<0.0001	ND	0.71±0.03	4.33±0.1	0.0085
Sulfate (mg/L)	6.68±1.0	26.98±0.1	3.51±0.1	<0.0001	40.49±0.5	22.91±1.3	58.11±1.4	0.0002
NH_3 –N	0.04±0.01	0.17±0.04	ND	0.0001	0.06±0.01	0.13±0.04	0.04±0.01	<0.0001
Nitrate (mg/L)	ND	ND	ND	NA	31.656±0.1	ND	ND	NA
Nitrite (mg/L)	ND	ND	ND	NA	ND	ND	4.25± 0.03	NA

ND - not-detected, NA – not applicable

influent and effluent samples. Statistically, the level of pH, dissolved oxygen, dissolved organic carbon; total dissolved solids, salinity and conductivity were significantly higher ($p<0.05$) in the studied influents of the WWTPs (Table 1).

Taxonomic Analysis of the Microbial Communities at the Treatment Plants

In total 5 611,755 to 13 602,600 reads were obtained after quality control for each of the six influent and effluent samples, they were assembled and the homologous sequences were removed and classified. Taxonomic classification showed that an ample fraction of about 30% reads in each metagenome had no matches to known organisms. A large fraction of the assembled reads showed dominance of *Bacteria* (69.7-95.2%), followed by small proportions of *Archaea* (0.13-1.17%), Eukaryota (0.8-12.45%), and Viruses (0.44-2.87%). The relative abundances of bacterial groups in the influent and effluent samples were analyzed at the phylum, class, and genus levels. At the phylum level the dominant groups were *Proteobacteria, Firmicutes*, *Bacteroidetes* and *Actinobacteria*. At the class level, the dominant classes were *Alphaproteobacteria, Actinobacteria, Bacilli*, and *Clostridia* respectively. The Archaea domain was dominated mainly with the *Crenarchaeota*, *Euryarchaeota*, *Korarchaeota* and *Thaumarchaeota* in both influent and effluent samples. They were mainly dominated by the *Opisthokonta, Fungi, Dikarya, Chlamydomonadales* and *Ostreococcus tauri* while the members of the *methanobacteria* and *methanococcus* class were observed to dominate the archaeal domain in both the influent and effluent samples of both plants. Mostly the *Oligohymenophorea, Amphibia, Streptophyta* class, dominated the eukaryotic domain and some unclassified classes were observed.

The viral taxonomic composition of all wastewater samples sequencing data (IFP, EFP, IDP, EDP, IPP) were analyzed using METAVIR online tool. The predominant viral reads across all samples belongs to the dsDNA viral group as presented in Fig. 1. The reads assigned to the dsDNA accounted for about 52% of all the viral sequences from EFP, IDP, EDP, IPP, and EPP, except for the IFP sample with about only 30%. The ssDNA reads accounted for 5- 24%, ssRNA 6.4-26.5%, and dsRNA 2.7-20.1%. As anticipated in both influent and effluent samples most of the detected sequences belong to the dsDNA viruses that are linked mostly to the bacteriophages.

Siphoviridae was the most abundant viral family and it was identified in both influent and effluent samples, this accounted for the largest percentage of the identified viromes. *Siphoviridae* are members of the dsDNA viruses whose natural hosts are bacteria and archaea. Fig. 2 represents the relative abundant and a taxonomic distribution of the viral families from each WWTPs. *Microviridae* family was also abundant in samples IFP (32%), EFP (18.1%). *Microviridae* is also a family of bacteria-infecting ssDNA viruses which are still being characterized. Similarly, *Baculoviridae* was also dominant in influent samples IPP (69.7%) and IDP (61.5%). *Baculoviridae* is another viral family where arthropods and invertebrates serve as their natural hosts. Apart from families affiliated with bacteriophages, other notable viral families identified were the viruses that infect plants such as *Partitiviridae* (2.6-18.3%),

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

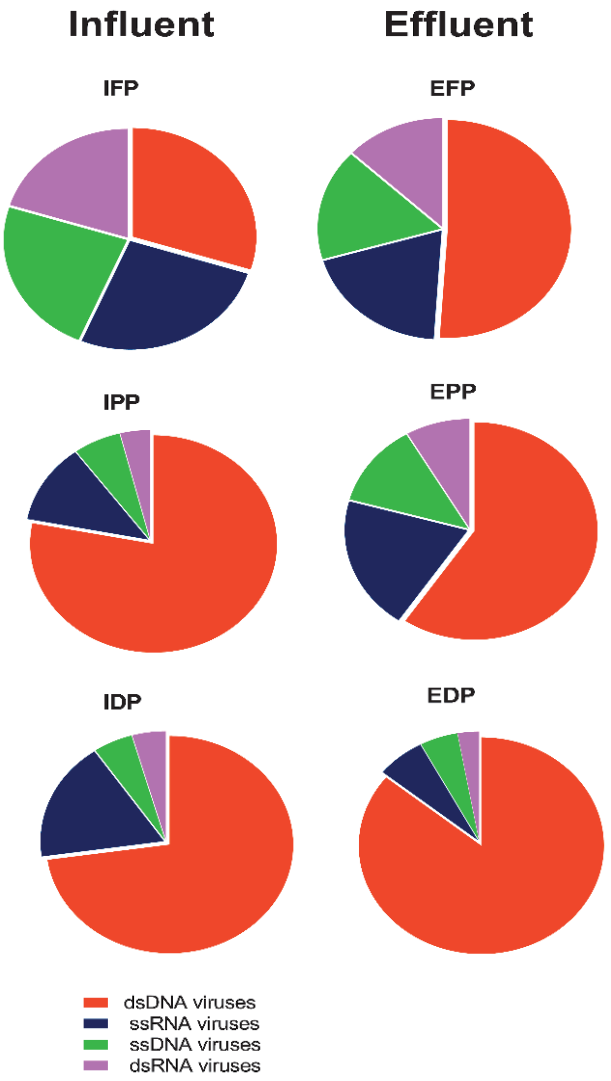

Fig. 1. Relative abundance of the viral taxonomic distribution across three WWTPS in Gauteng Province, South Africa.

Virgaviridae (0.5-13.6%), and *Tymoviridae* (0.9-9.5%). Sequences that are often associated to some common enteric viruses such as *Astroviridae, Retroviridae*, and *Coronaviridae* were detected. However, viral families with <1% sequences were grouped into minor families. Viruses infecting bacterial genera containing pathogens, including *Salmonella* (2.5%), *Pseudomonas* (4.6%), *Burkholderia* (1.8%), *Klebsiella* (1.3%), *Vibrio* (5.0%), and some viral families associated with the infection of cyanobacteria were also detected.

At the genus level, most classified viral sequences belonged to *Lambdalikevirus* of the order Caudovirales, in the family *Siphoviridae*. This was followed by the *Betabaculovirus*, *Tobamovirus, Microvirus, Alphapartitivirus, and Betapartitivirus*. Other observed dominant genera included the *Tymovirus, Flavivirus, Hepacivirus, N15-likevirus* and *T4-like viruses* of the *Myoviridae* family as shown in Fig. 3. This therefore indicates that phages are playing a major diversity role in the studied wastewater treatment plants. The diversity of the viromes in influent and effluent samples was characterized using Bray Curtis distances. Their significance was tested and the output was visualized using PCoA analysis, with a significant difference of (p<0.05). To explore the compositional similarities among the viral communities in the three treatment plants, a PCoA of weighted UniFrac distances was constructed from the OTU count data. PCoA bi-plot analyses showed a clear segregation of the viral communities by WWTPS (Fig. 4). Samples EDP communities clustered along the first axis, while the viral communities from the IPP, IDP samples were distributed on the second quadrant. Samples EPP, EFP and IFP sample communities were distinctively distributed over the third quadrant of the plot.

Functional Annotation and Metabolic Analysis

The metabolic profile of the six metagenomes was explored using MG-RAST. The functional role of predicted genes from the viral assemblies was determined by performing a BLASTX hit against the functionally annotated SEED and COG database with e-value $<10^{-5}$. In total 2.49- 13.60% of the sequences were functionally classified. The most represented categories were related to the metabolism of clustering-based subsystem, reads associated with protein and carbohydrates metabolism were the second most observed. Amino acid, phage, prophage, virulence, cofactors, vitamins, RNA/DNA metabolism, and nucleosides/nucleotides were the third most identified categories as shown in Fig. 5. Functional distribution at the subsystem hierarchy showed that about 75% of sequences in both the influent and effluents were predicted as phages or prophages in the dataset respectively. The sequences also demonstrated the occurrence of virulence, disease and defense factors in the influent and effluent samples occupying about 12-40% of the functional hits.

Viral Pathogens

The most frequently amplified viral sequences belonged to the viral families infecting invertebrates, vertebrates, plant viruses and viruses found in diverse environmental samples. Of the annotated reads, about 30% were identified plant viral pathogens, while the remaining was 70% was regarded as either animal or human origin. The most abundant plant viral pathogen family observed in this study was the *Partitiviridae* family with abundance of 2.6%-18.3% in both influent and effluent samples as shown in Fig. 6, the main natural hosts of this viral family are fungi and plants. Plant virus signature belonging to the family *Virgaviridae*, were present in all wastewater samples from (0.9-9.6%). Other highly represented plant viral families were the families *Tymoviridae* occurring between

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

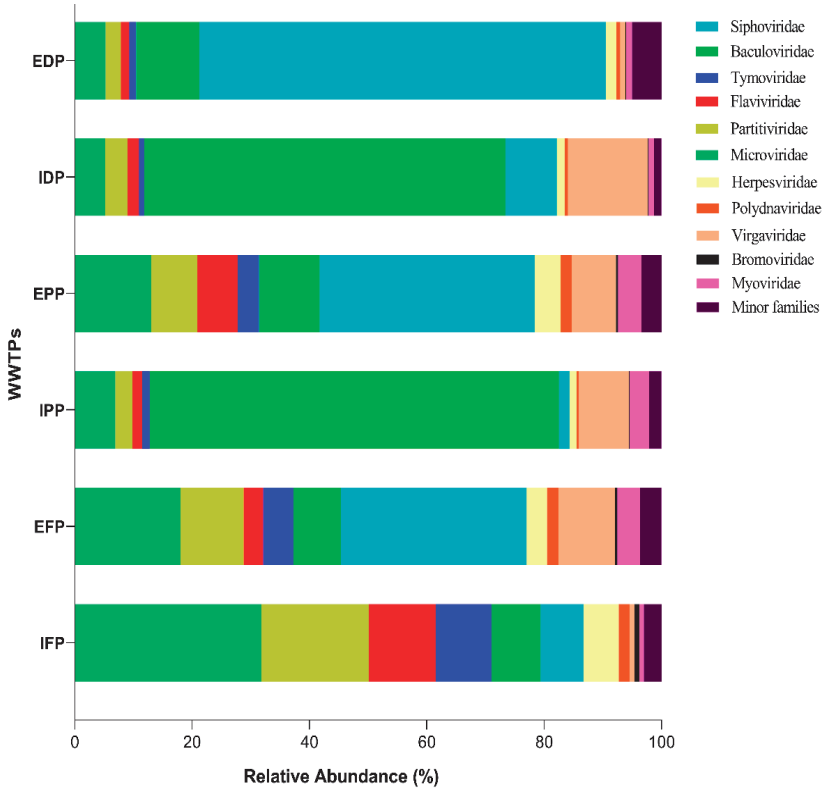

Fig. 2. Relative abundance of the predominant viral family exhibited at WWTPs based on METAVIR.

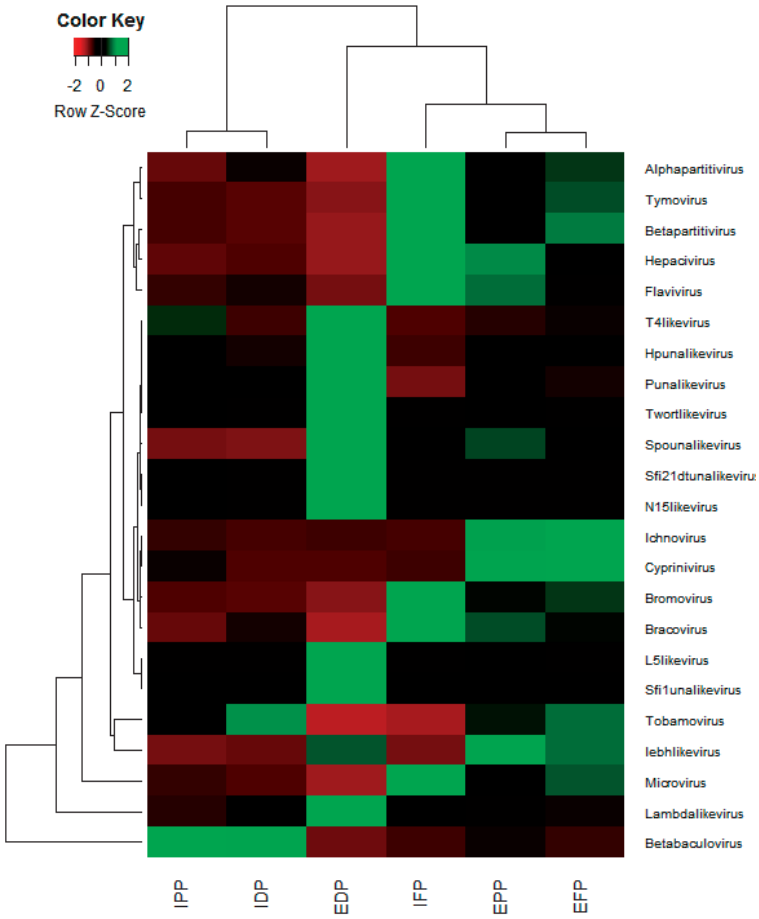

Fig. 3. Heatmap of the most abundant genera in the WWTPS indicating the viral richness at genus level per sample. The heatmap color indicates the relative abundances per sample using the lowest-common ancestor algorithm.

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

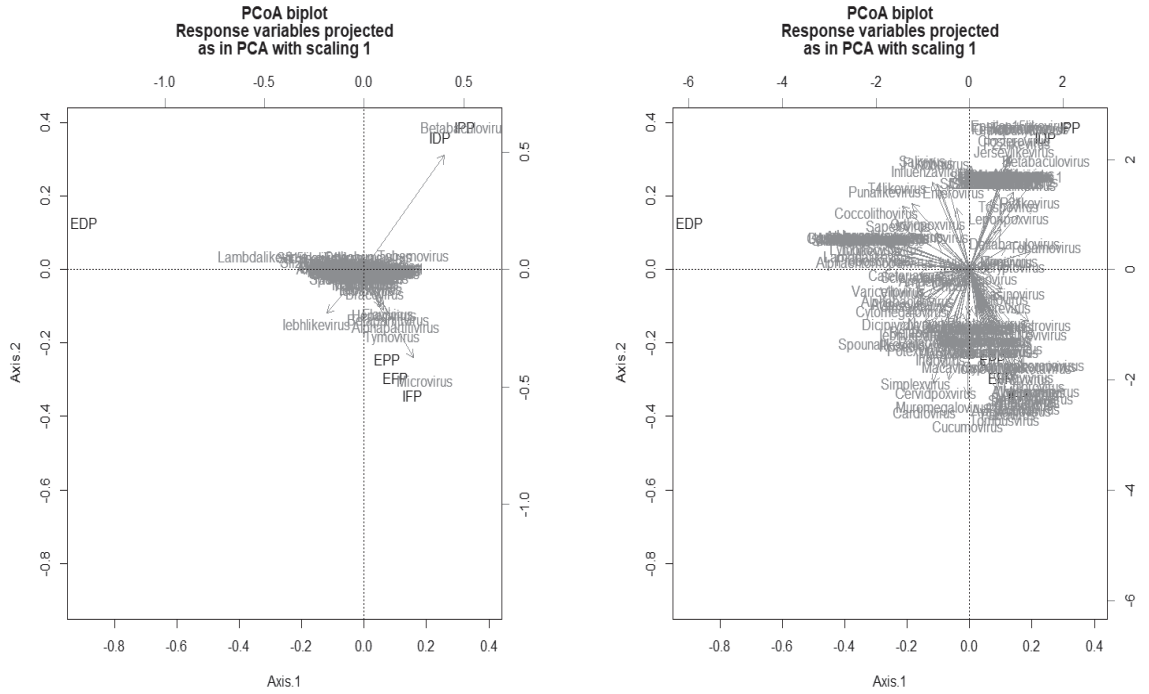

Fig. 4. Beta-diversity based on Bray Curtis distance, shown in Principal coordinates analysis (PCoA) plot of the viral communities displaying clusters of viral communities based on weighted UniFrac distance matrix of species richness.

0.9-9.5% across samples, *Tombusviridae* and *Potyviridae* occurring mostly in sample IFP at comparable level. This is regarded of great importance mainly due to the fact that during agricultural activities within the agriculture sector as some plant viruses are able infect plant produce through the use of contaminated water sources during cultivation and irrigation activities (Fig. 6a). Overall, 20 families of viruses were identified that affect insects, birds, mammals and humans. The relative abundance and occurrence of human and animal viral pathogens population are presented in Fig. 6b). The viral families dominating at the three WWTPs with the majority of annotated contigs belonged to the family of *Baculoviride* (insects). They were detected mostly in samples influent samples of IDP (61.5%), and IPP (69.7%) respectively. The *Flaviviridae* family was dominant in sample IFP (11.4%), other frequently detected human pathogenic viruses found

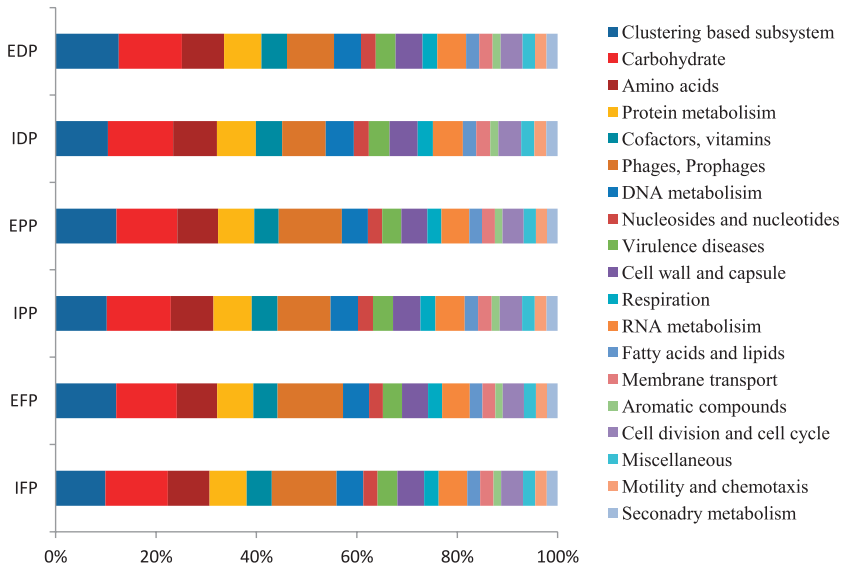

Fig. 5. Relative abundances of sequences assigned to each metabolic subsystem by MG-RAST. The metabolic categorization is based on the sequences best Blast hits in the SEED database curated subsystems (*E*-value $<1e^{-05}$).

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

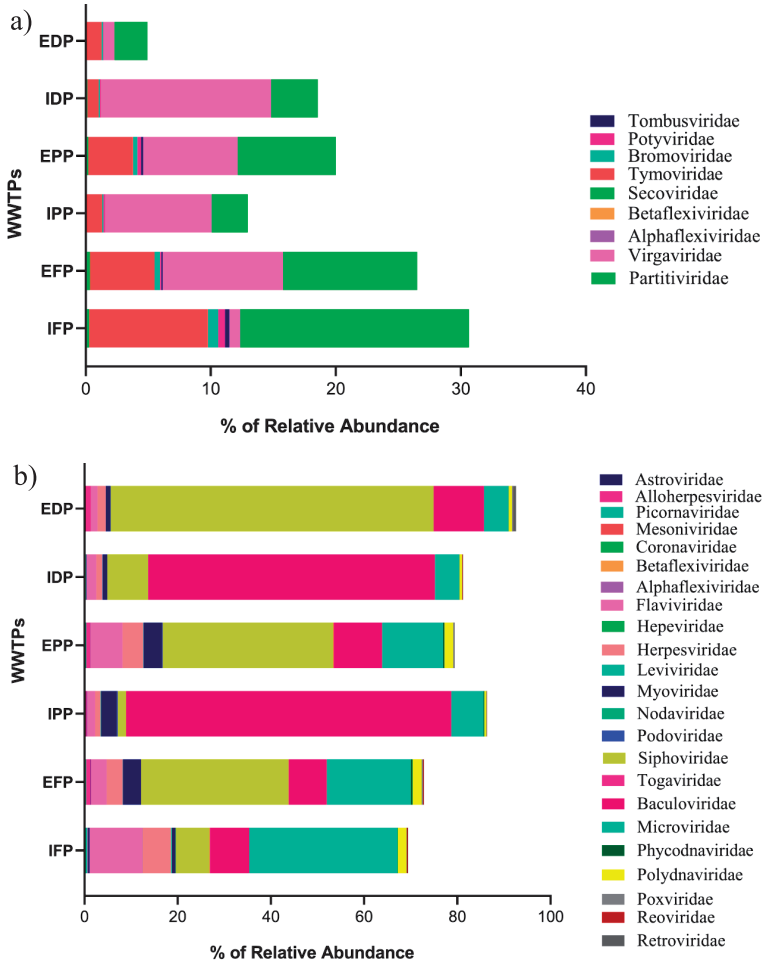

Fig. 6. Diversity of potential a) plant pathogenic and b) human and animal pathogenic viruses and their relative abundance at family level in the WWTPs.

in the influent and effluent samples were *Astroviridae, Retroviridae, Coronavivirida* and *Picornaviridae.*

Wastewater samples are characterized by differentiated physical-chemical properties (Table 1), the largest differences in the physiochemical properties were observed in the case of conductivity with high values been detected in all samples, which could be attributed to the presence of suspended and dissolved ions in the water samples. WWTPs play a trivial role in the aquatic ecosystem; pH of the wastewater samples varies significantly at each sample points and ranged from 6.64 to 8.12 which falls within the World Health Organization standard specification and the water quality range. High phosphate values from the plants were observed which ranged from 5.07-23.29 mg/L, this indicates the heavy presence of industrial, cleaning and farming activities in the surroundings. Nitrate was only detected in effluent sample of EDP with a value of 31.656 mg/L; this gives a clue about the different level of micronutrients in the wastewater and hence their capability to support plant growth.

To understand the total viral community at the wastewater treatment plants (influent/effluents), the microbial analysis of the communities showed the dominance of about 70% bacteria domain while the remaining percentage was shared among other domains. The cluster analysis showed that the taxonomic composition of the influent and effluent samples were very much similar: they all had a bacterial community composition enriched with *Proteobacteria, Bacteroidetes, Firmicutes,* and *Enterobacteriaceae* which are regarded as major concerns for human health as reported in previous studies [25]. The sequence results revealed phylum *Proteobacteria* (36.14%) as the dominant community followed by the *Firmicutes* (34%), *Bacteroidetes* (11.71%) and *Enterobacteriaceae* (8.5%). Within the *Proteobacteria* phylum, *Alphaproteobacteria* was the most abundant class across the wastewater samples. Other noticeable classes included *Bacilli, Actinobacteria*, *Clostridia, Gammaproteobacteria* and *Flavobacteria* in all samples.

Other major domain in the influent and effluent samples were belong to Archaea which were dominated by *Crenarchaeota*, *Euryarchaeota*, and *Thaumarchaeota,* which are known to strictly preforming functions such reducing carbon dioxide and other bacterial waste product into methane within the treatment plant. The dominant Archaea detected within

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

the WWTP influent and effluents belongs to the genus *Methanobacteriales, Desulfurococcales, Desulfovibrio* and *Methanococcales, Methanosarcinales,* which are known to strictly preforming functions such reducing carbon dioxide to methanogens. *Desulfovibrio sulfuricans* are abundance in WWTPs is vital because they produce acetate that could be consumed by methanogens especially within the microbial population. A number of *Desulfovibrio* species have been reported to produce acetate, CO_2 and H_2 in co-occurrence with hydrogenotrophic methanogens during limited sulfate conditions.

A total of 43 different viral families were identified in this study, with about 12 families dominant in each plant. Studies have indicated that WWTPs obtain water from various point and non-point sources, thereby making them a creek for the collection of many viral families that infect a large range of hosts [3, 47-48]. The largest viral genomes observed in this study were dominated by the Caudovirales order, with families such as *Siphoviridae, Microviridae* and *Myoviridae* conspicuous. *Siphoviridae* were the most abundant in all three plants effluents making up to 7.3-69.3%, *Myoviridae* abundance was between 1.5-4% and *Microviridae* 5-32% across all samples. These viral families belonged to the group of tailed bacteriophages which are known to infect a wide range of bacterial cells and are hosts specific [45]. The relative abundance of phages belonging to the *Siphoviridae* family in the WWTPs untreated sewage and treated effluents is consistent with other viral metagenomic studies in literature suggesting that *Siphoviridae* are potential indicators of fecal pollution [24, 46]. Also, the dominance of *Siphoviridae* in WWTP effluents put forward that members of this viral family may be resistant to chlorine treatment [46]. A number of studies have also highlighted that phages from the *Siphoviridae* family are generally more resistant to chlorination than other specific phage families [46]. Based on these observations, it is suggested that members of the *Siphoviridae* family could be further explored as potential viral indicators as they denote an abundant group of viruses in sewage that are fairly resistant to treatment at WWTPs.

The main dominating ssDNA virus in the samples was the *Baculoviridae* family and they are regarded as insect specific viruses, they were mostly abundant in the influent sample of IDP. These water related family viruses have water related insects as their host such as mosquitoes. Besides, ssDNA viruses are small and similar to RNA viruses with genomes ranging from about 2-6 kb, some of these viruses are regarded as common agricultural pests. Another main group of viruses that dominated in the samples were the ssRNA viruses that infect the eukaryotic cells with examples such as *Tombusviridae, Virgaviridae,* and *Alphaflexiviridae*.

At the genus level, the most dominant genera belonged to *Betabaculovirus* in the influent samples with 73% in IPP and 65% in IDP. The *Microvirus* genus was dominant in both the influent and effluent samples of IFP 36%, EFP 25%. The most common phages found in wastewater samples were *Lambdalikevirus* which belongs to the order Caudovirales, in the family *Siphoviridae*. They are known to infect the bacterial species enterobacteria and they were found with high occurrence only in the effluent of EDP at 57%. Others found were the *Tobamovirus,* 14% in IDP, *Alphapartitivirus* 10% in IFP, and *Betapartitivirus,* between 1-6% in all samples. Other observed genera in both influent and effluent samples included the, *Tymovirus* 1-11%, *Flavivirus* 1.2-7.9%, *Hepacivirus,* N15-likevirus and T4-like viruses (1-5%).

The diversity of assigned viral sequences was analyzed based on number of unique viral species (species richness), diversity was higher in the effluent than in the influent samples of the three WWTPs. The principal coordinate's analysis (PCoA) based on the Bray-Curtis distance metric showed that viral communities clustered significantly by individual WWTP. The three WWTPs exhibited distinct viral communities when compared to one another and the detected viral species also varied slightly within each WWTP across sampling sites. This also indicates that effluent samples EPP and EFP are introducing some viral species back into the effluents that is been released into the environment. It is very important to identify the virus diversity as the lack of knowledge on characteristics of the viral community could hinder vital sources of emerging pathogens as well as the exchange of gene amongst viruses [2]. Notably, number of sequences obtained in this study showed no sequence relation to any known sequences in the viral databases which may likely be as a result of either been uncharacterized viruses or emerging viral species. Using the subsystems approach, there were no conspicuous changes observed for each functional category among the different sample's metabolism [48]. Among the metabolism-related category, carbohydrate transport and metabolism were the most predominant (9.5-12%), amino acid transport and metabolism (7-8.6%), and lipid transport and metabolism (2-3.8%) indicating bacterial contamination. The DNA/RNA metabolism was between (5-7%), cell division and cell cycle functional category showed lesser values after DNase treatment in all samples (4.1%), while the clustering-based subsystems (unknown functions) were 13.6%. The data also demonstrate the occurrence of virulence, disease and defense factors in the WWTP samples occupying 10-38% of functional traits. Analysis of the virulence (disease and defense functional category) revealed a higher resistance to antibiotics all samples analysed which is a major cause for concern to public health. Excitingly, the remaining viral reads that were unassigned may likely indicate the presence of novel genetic viral diversity within the studied plants.

The occurrence of viruses with known pathogenicity was also investigated at the WWTPs, with major focus

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

on plant, animal/human viruses and their economic importance. The most dominant plant viral genome mostly detected in the analyzed wastewater samples in this study was the *Partitiviridae* viral family. They were observed to be persistent even after treatment in all samples, *Partitiviridae* natural hosts are plants, fungi, or protists [49]. Another conspicuous viral family detected was the *Virgaviridae* family, also present in influent and effluent samples at comparable levels which is consistent with some previous studies [29, 31]. In general, among the *Virgaviridae* family a high abundance of plant viruses which include the Tobamovirus genus, Pepper mild mottle virus (PMMoV) and Tobacco mild green mosaic virus (TMGMV) species were detected. PMMoV is a pathogen of pepper; their main source is through the consumption of pepper food products and then excreted in human faeces. Although some studies have suggested that PMMoV is a prospective indicator of human fecal contamination of environmental waters [18, 50-52]. This is consistent with studies that have highlighted the contamination of domestic water sources with PMMoV in some countries [49, 50-54]. Further, highly represented sequences belonged to the *Tymoviridae* families which were dominant in samples IFP, EFP and EPP between (5-7%); and *Tombusviridae* was in all samples except sample EDP. These observations are also consistent with previous studies that WWTPs contains different plant viruses [2]. Other families of interest which are of importance to agriculturist and farmers were the *Alphaflexiviridae, Bromoviridae, Potyviridae,* and *Leviviridae* which contains a quite amount of species that destroys the leaves of some plants. This finding exposes how WWTPs and anthropogenic environments allow the movement and spread of plant diseases into the environment [4]. However, in some places where the sludge produced during the treatment process at WWTPs is often used as fertilizer, if the sludge is not adequately disinfected, likely pathogens present might be re-introduced back into the food produce through fertilizer application or consumption.

Some other pathogenic viruses identified by metagenomics were the *Herpesviridae, Flaviviridae, Poxviridae, Picornaviridae, Alloherpesviridae* and *Togaviridae* with occurrence of about 0.1-3.8% in all samples. Some of these viruses have mammals, birds and insects as their hosts. *Herpesviridae* is a large viral family of concern because it can cause diseases in humans and animals, members such as *Herpesviruses* having quite a handful number of species [9, 55]. *Herpesvirus* has been regarded as the leading cause of human viral disease, alongside influenza and cold viruses. Within the *Flaviviridae* family humans and other mammals serve as natural hosts, they are primarily spread through arthropod vectors such ticks and mosquitoes. On the other hand, some common signature enteric pathogens such as *Astroviridae, Reoviridae,* and *Coronaviridae* were also identified but in lower quantities. Sequences from the Reoviridae viral family was observed and represented mostly by the genus *Rotavirus* in all the samples. *Rotavirus* are classified among the enteric viruses transmitted through the fecal-oral route, it is a very contagious virus that causes diarrhea and acute gastroenteritis worldwide in humans [1, 56-57]. Studies have shown that they can persist in the environment for a long time and they are highly resistant to treatment processes at WWTPs [58]. Another enteric virus detected in the samples were *Astrovviridae* viral family and they were represented by the genus *Astrovirus,* they are known to cause moderate to mild diarrhoeal illnesses [59-61]. Another virus of concern that was detected in all the samples was the *Coronoviridae* viral family; these are viral species that are primarily capable of causing respiratory illness and common cold infections especially in people with compromised health issues. A more comprehensive studies need to be carried out to monitor the occurrence and persistence of this viral species transmission rate within the community. Overall, a number of unclassified and unassigned ssRNA environmental viruses which were not shown were also widely distributed in all the samples.

Conclusions

In conclusion, the results of this study serve to expand our view on the abundance and diversity of the different viral communities in influent and effluent samples of three different WWTPs in South Africa using high throughput sequencing technology. The metagenomic analyses unveiled that the WWTPs studied were are highly dominated with Caudovirales consisting of *Siphoviridae*, *Microviridae* and *Myoviridae* viral families. *Viviridae* and *Partitiviridae* viral families were the major plant infecting virus dominating the plants with different species. This helps buttress the fact that that high throughput sequencing can provide a comprehensive study into the presence of plant viruses, reflecting the type of food crops or produce that is been consumed by residents and animals within the sampling locations. Other viral families detected include *Baculoviride*, *Flaviviridae*, *Reoviviridae, Astroviridae* and *Coronaviridae.* These suggest that current regulations for pathogens in WWTPs that focus on fecal coliform indicators do not capture the full degree of pathogenic diversity to which the public may be exposed. These findings also demonstrate the need for developing an improved suite of indicator pathogens protocols and lists. Further research should focus on seasonal monitoring of the final effluent and the activated sludge samples to determine which viral families can still persist and deserved more concerns. In spite of several fundamental setback associated with viral analysis through metagenomics, there are still more advantages over existing techniques as viral metagenomics analysis provides a broader spectrum

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

of uncovering the diversity and unknown viral genes that may be potential pathogens.

Acknowledgements

The authors are thankful to the University of South Africa (UNISA) for supporting this research. The financial assistance of the National Research Foundation (grant SFH170629246463) towards this research is hereby acknowledged. The authors also would like to thank CHPC, Pretoria for proving high computing facility for sequence analysis.

Conflicts of Interests

The authors declare that there is no conflict of interests.

References

1. OSUOLALE O., OKOH A. Human enteric bacteria and viruses in five wastewater treatment plants in the Eastern Cape, South Africa. J of Infection and Public Health, **10** (5), 541, **2017**.
2. CANTALUPO P.G., CALGUA B., ZHAO G., HUNDESA A., WIER A.D., KATZ J.P., PIPAS J.M. Raw Sewage Harbors Diverse Viral Populations. mBio, **2** (5), 1, **2011**.
3. TAMAKI H., ZHANG R., ANGLY F.E., NAKAMURA S., HONG P.Y., YASUNAGA T., LIU W.T. Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. Environ Microb, **14** (2), 441, **2012**.
4. ROSARIO K., NILSSON C., LIM Y.W., RUAN Y., BREITBART M. Metagenomic analysis of viruses in reclaimed water. Environ Microb, **11** (11), 2806, **2009**.
5. MANAIA C.M., ROCHA J., SCACCIA N., MARANO R., RADU E., BIANCULLO F., NUNES O.C. Antibiotic resistance in wastewater treatment plants: Tackling the black box. Environ Internat **115** (March), 312, **2018**.
6. CHE Y., XIA Y., LIU L., LI A.D., YANG Y., ZHANG T. Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. *Microbiome*, **7** (1), 1, **2019**.
7. LIN J., GANESH A. Water quality indicators: Bacteria, coliphages, enteric viruses. Int J. of Env Health Research, **23** (6), 484, **2013**.
8. PETROVICH M.L., BEN MAAMAR S., HARTMANN E.M., MURPHY B.T., PORETSKY R.S., WELLS G.F. Viral composition and context in metagenomes from biofilm and suspended growth municipal wastewater treatment plants. Micro Biotech, **12** (6), 1324, **2019**.
9. O'BRIEN E., NAKYAZZE J., WU H., KIWANUKA N., CUNNINGHAM W., KANEENE J.B., XAGORARAKI I. Viral diversity and abundance in polluted waters in Kampala, Uganda. Wat Res, **127**, 41, **2017**.
10. OSUNMAKINDE C.O., SELVARAJAN R., SIBANDA T., MAMBA B.B., MSAGATI T.A.M. Overview of trends in the application of metagenomic techniques in the analysis of human enteric viral diversity in Africa's environmental regimes. Viruses, **10** (8), **2018**.
11. KIULIA N.M., NETSHIKWETA R., PAGE N.A., VAN ZYL W.B., KIRAITHE M.M., NYACHIEO A., TAYLOR M.B. The detection of enteric viruses in selected urban and rural river water and sewage in Kenya, with special reference to rotaviruses. J. of App Microbiology, **109** (3), 818, **2010**.
12. GIBSON K.E. Viral pathogens in water: Occurrence, public health impact, and available control strategies. Curr Opinion in Virology **4**, 50, **2014**.
13. AW T.G., HOWE A., ROSE J.B. Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater. J. of Virological Methods, **210**, 15, **2014**.
14. MONTAZERI N., GOETTERT D., ACHBERGER E. C., JOHNSON C.N., PRINYAWIWATKUL W., JANES M.E. Pathogenic enteric viruses and microbial indicators during secondary treatment of municipal wastewater. Applied and Environmental Microbiology, **81** (18), 6436, **2015**.
15. FONG T.T., LIPP E.K. Enteric viruses of humans and animals in aquatic environments: Health risks, detection, and potential water quality assessment tools. Microbiology and Molecular Biology Reviews, **69** (2), 357, **2005**.
16. BOFILL-MAS S., RUSIÑOL M., FERNANDEZ-CASSI X., CARRATALÀ A., HUNDESA A., GIRONES R. Quantification of Human and Animal Viruses to Environmental Samples, **2013**, **2013**.
17. IKNER L.A., GERBA C.P., BRIGHT K.R. Concentration and Recovery of Viruses from Water: A Comprehensive Review. Food and Environmental Virology, **4** (2), 41, **2012**.
18. HAMZA I.A., JURZIK L., ÜBERLA K., WILHELM M. Methods to detect infectious human enteric viruses in environmental water samples. International Journal of Hygiene and Environmental Health, **214** (6), 424, **2011**.
19. HARAMOTO E., KITAJIMA M., HATA A., TORREY J.R., MASAGO Y., SANO D., KATAYAMA H. A review on recent progress in the detection methods and prevalence of human enteric viruses in water. Water Research, **135**, 168, **2018**.
20. MCMINN B.R., ASHBOLT N.J., KORAJKIC. Bacteriophages as indicators of faecal pollution and enteric virus removal. Lett in Appl Microbiology, **65** (1), 11, **2017**.
21. MENDEZ J., MONTEMAYOR M., PAYA A. Microbial indicators and pathogens: Removal, relationships and predictive capabilities in water reclamation facilities Microbial indicators and pathogens: Removal, relationships and predictive capabilities in water reclamation facilities, (August), **2008**.
22. HOT D., LEGEAY O., JACQUES J., GANTZER C., CAUDRELIER Y., GUYARD K. Detection of somatic phages, infectious enteroviruses and enterovirus genomes as indicators of human enteric viral pollution in surface water, **37**, 4703, **2003**.
23. PRONK M., GARCIA H., BOERSMA A., BARRIOS-HERN M.L., BRDJANOVIC D., LOOSDRECHT M.C.M. VAN, HOOIJMANS C.M. Removal of bacterial and viral indicator organisms in full-scale aerobic granular sludge and conventional activated sludge sYSTEMS, **6**, **2020**.
24. ADRIAENSSENS E.M., FARKAS K., HARRISON C., JONES D.L., ALLISON H.E., MCCARTHY A.J. Viromic Analysis of Wastewater Input to a River Catchment Reveals a Diverse Assemblage of RNA Viruses. mSystems, **3** (3), 1, **2018**.
25. OSUNMAKINDE C.O., SELVARAJAN R., MAMBA B.B., MSAGATI T.A.M. Profiling bacterial diversity and potential pathogens in wastewater treatment plants using

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

high-throughput sequencing analysis. Microorganisms, **7** (11), **2019**.
26. LELAND D.S., GINOCCHIO C.C. Role of cell culture for virus detection in the age of technology. Clinical Microbiology Reviews, **20** (1), 49, **2007**.
27. GERBA C.P., BETANCOURT W.Q. Assessing the Occurrence of Waterborne Viruses in Reuse Systems : Analytical Limits and Needs. **2019**.
28. RODRÍGUEZ R.A., PEPPER I.L.,GERBA C.P. Application of PCR-based methods to assess the infectivity of enteric viruses in environmental samples. Applied and Environmental Microbiology, **75** (2), 297, **2009**.
29. BIBBY K., VIAU E., PECCIA J. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. Letters in Applied Microbiology, **52** (4), 386, **2011**.
30. BIBBY K., VIAU E., PECCIA J. NIH Public Access, **52** (4), 386, **2012**.
31. BIBBY K. Metagenomic identification of viral pathogens. Trends in Biotechnology, **31** (5), 275, **2013**.
32. WU Q., DING S.W., ZHANG Y., ZHU S. Identification of Viruses and Viroids by Next-Generation Sequencing and Homology-Dependent and Homology-Independent Algorithms. Ann Rev of Phytopathology, **53** (1), 150605182533006, **2014**.
33. ADAMS R.I., BHANGAR S., DANNEMILLER K.C., EISEN J.A., FIERER N., GILBERT J.A., BIBBY K. Ten questions concerning the microbiomes of buildings. Building and Environment, **109**, 224, **2016**.
34. BIBBY K. Metagenomics and the development of viral water quality tools. npj Clean Water, (January), 1, **2019**.
35. PEARSON V.M., CAUDLE S.B., ROKYTA D.R. Viral recombination blurs taxonomic lines: Examination of single-stranded DNA viruses in a wastewater treatment plant. PeerJ, **2016** (10), 1, **2016**.
36. BREITWIESER F.P., LU J., SALZBERG S.L. A review of methods and databases for metagenomic classification and assembly. Briefings in Bioinformatics, (September), 1, **2017**.
37. IKNER L.A., SOTO-BELTRAN M., BRIGHT K.R. New method using a positively charged microporous filter and ultrafiltration for concentration of viruses from tap water. Applied and Environmental Microbiology, **77** (10), 3500, **2011**.
38. BOLGER A.M., LOHSE M., USADEL B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, **30** (15), 2114, **2014**.
39. PENG Y., LEUNG H.C.M., YIU S.M., CHIN F.Y.L. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics, **28** (11), 1420, **2012**.
40. ROUX S., FAUBLADIER M., MAHUL A., PAULHE N., BERNARD A., DEBROAS D., VALENCIA A. Metavir: a web server dedicated to virome analysis, **27** (21), 3074, **2011**.
41. EDWARDS R.A., SCHMIEDER R., ANGLY F.E., WILLNER D., PRIETO-DAVO A., VEGA-THURBER R., WEGLEY L. The GAAS Metagenomic Tool and Its Estimations of Viral and Microbial Average Genome Size in Four Major Biomes, **5** (12), **2009**.
42. WOOD D.E., SALZBERG S.L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biology, **15** (3), **2014**.
43. NAIDOO S., OLANIRAN A.O. Treated wastewater effluent as a source of microbial pollution of surface water resources. Intern J of Environ Res and Public Health, **11** (1), 249, **2013**.
44. FEI T., NG F., MARINE R., WANG C., SIMMONDS P., KAPUSINSZKY B., PHYLOGE N. High Variety of Known and New RNA and DNA Viruses of Diverse Origins in Untreated Sewage, **86** (22), 12161, **2012**.
45. HAYES S., MAHONY J., NAUTA A., VAN SINDEREN D. Metagenomic approaches to assess bacteriophages in various environmental niches. Viruses, **9** (6), 1, **2017**.
46. OVERBEEK R., BEGLEY T., BUTLER R.M., CHOUDHURI J.V., CHUANG H., COHOON M., VONSTEIN V. The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes. Nucleic acids research, **33** (17), 5691, **2005**.
47. ROOSSINCK M.J. Tansley review Evolutionary and ecological links between plant and fungal viruses. **2018**.
48. HARAMOTO E., KITAJIMA M., KISHIDA N., HARAMOTO E., KITAJIMA M., KISHIDA N., ASAMI M. Occurrence of Pepper Mild Mottle Virus in Drinking Water Sources in Japan. Appl and Environ Microbiology, **79** (23), 7413, **2013**.
49. SYMONDS E.M., GRIFFIN D.W., BREITBART M. Eukaryotic viruses in wastewater samples from the United States. Appl and Environ Microbiology, **75** (5), 1402, **2009**.
50. SYMONDS E.M., NGUYEN K.H., HARWOOD V.J., BREITBART M. Pepper mild mottle virus: A plant pathogen with a greater purpose in (waste)water treatment development and public health management. Water Research. **2018**.
51. KITAJIMA M., SASSI H.P., TORREY J.R. Pepper mild mottle virus as a water quality indicator. npj lCean Water, **1** (August), 19, **2018**.
52. BRIEN C.A.O., MCLEAN B.J., COLMANT A.M.G., HARRISON J.J., HALL-MENDELIN S., HURK A.F., VAN DEN, HOBSON-PETERS J. Discovery and Characterisation of Castlerea Virus, a New Species of Negevirus Isolated in Australia. Evolutionary Bioinformatics, 1, **2017**.
53. OSUOLALE O., OKOH A. Incidence of human adenoviruses and Hepatitis A virus in the final effluent of selected wastewater treatment plants in Eastern Cape Province, South Africa. Virology Journal, **12** (1), 98, **2015**.
54. ZHOU N., LV D., WANG S., LIN X., BI Z., WANG H., WANG P. Continuous detection and genetic diversity of human rotavirus A in sewage in eastern. Virology Journal, 1, **2016**.
55. HARAMOTO E., KATAYAMA H., UTAGAWA E., OHGAKI S. Development of sample storage methods for detecting enteric viruses in environmental water. J of Virol Methods, **151** (1), 1, **2008**.
56. MARTÍNEZ M.A., DE LOS DOLORES SOTO-DEL RÍO M., GUTIÉRREZ R.M., CHIU C.Y., GRENINGER A.L., CONTRERAS J.F., ISA P. DNA microarray for detection of gastrointestinal viruses. J of Clin Microbiology, **53** (1), 136, **2015**.
57. HELLMÉR M., PAXÉUS N., MAGNIUS L., ENACHE L., ARNHOLM B., JOHANSSON A., NORDER H. Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. Appl and Env Microbiology, **80** (21), 6771, **2014**.