nature portfolio

Peer Review File

Unveiling the global urban virome through wastewater metagenomics

Corresponding Author: Dr Miranda de Graaf

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The research led by Worp et al., analyzes DNA and RNA viromes from different parts (62 cities of the world. They use shotgun metagenomics combined with target enrichment viruses using GastroCap probe. The manuscript is overall well written. Wastewater surveillance mostly targets bacterial population. Studies conducted with viruses are less common mainly due incomplete or incorrect information on databases. I did find some gaps and major comments in their methodology. There is no prophage analysis. There are no clear distinction among RNA and DNA viruses in terms of discussion and results. I mean that makes difficult for the reader to point out patterns. Results are presented as "total/whole" viruses, maybe a different pattern could be observed if DNA and RNA viruses are presented separately. Especially when PCA are used to demonstrate variance across the different dataset here presented.

Line 113: Is this value normal for viruses. What were the remaining reads classified as? Bacteria, human reads, microeukaryotes?. I had to see about the distribution in one of the figures.

Lines 171-172: Australia compared to other continents is not interconnected to any other one mass of land. Something that may be discussed/highlighted in the manuscript.

Line 304: Was the viral DNA fraction (aka phages) in this study screened for ARGs or prophages. That may provide valuable complementary information.

Line 340: Figure? or point out the months. Any particular parameter that could have driven this pattern.

Line 392: Value should be expressed in gravities instead of rpm.

Line 393: In addition 0.2 um filter could have been used. It is usual to get 16S rRNA gene or material still passs thru 0.45 um filters. While I acknowledge the use of OmniCleave there is no percentage of recovery or the efficiency of this commercial product to remove free DNA that could pose as contamination in the "viral" fraction. In addition, I did not see any experiment regarding a mock viral community that could have helped to present recovery rate plus sequencing control.

(Remarks on code availability)
The code is presented, but no data are currently not available. Thus, I could not tried the code presented by the authors.

Reviewer #2

(Remarks to the Author)
The study describes the detection of viral species prevalent in wastewater obtained from 62 cities located in six different continents. The study primarily used two approaches: shotgun metagenomic approach and a targeted approach to capture enteric viruses. The study revealed the vast diversity of viruses that infect a wide range of hosts spanning all major

kingdoms. As expected, the capture approach enriched the targeted viral families. Interestingly, the presence of certain groups of viruses were temporally correlated in some cases and in others clearly associated with a geographical location. This was further corroborated using PCA analysis revealing unique signatures associated with locations (cities) and the time of collection (seasonal). The study identified distinct viral diversity patterns and abundant viruses that influenced viral composition, which is relevant in plant, animal, and human health. The detection of plant viruses, especially those that appear to be emerging viruses, is important for agriculture and food security. Viruses detected only in wastewater indicated the inability of clinical surveillance in capturing the full diversity of human viruses. Wastewater monitoring also allowed tracking of enteric viruses such as polio, which is of immense clinical importance. The study also highlighted the presence of a large proportion of unassigned viral taxa, which can lead to the discovery of novel or emerging viruses. This study is an important step in understanding the application of wastewater-based monitoring in informing human, animal and even plant health.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
This paper sought to understand the potential for wastewater analysis as a method of disease and virus surveillance through a combination of shotgun metagenomic sequencing and GastroCap sequence probes. This approach addresses the need for better disease surveillance within cities and utilizes an approach that can be both cost effective and easy to implement. Samples were collected from a variety of geographical locations across multiple seasonal time frames, highlighting the consistency of the global viral community within wastewater environments. Additionally, shorter time frame sampling was able to provide insight into the more fine-tuned viral profile within cities, showing distinct differences between the families of virus present in different areas. The diversity of reported viral families and genera covers many well-defined groups in addition to several potentially novel viruses, making this of particular interest to virologists and molecular epidemiologists. The potential for this type of surveillance to improve public health is significant, as several of the collected samples across the globe were able to accurately predict specific outbreaks multiple months before they were first reported. Comments:

The viral metagenomics analysis performed is very thorough and they very rigorously characterize viral diversity.

Minor Notes:

1. While there is an amount of variety in the geography of these samples, I would argue that the lack of data for South and Central America impacts the meaning of your “global” and South American-centric conclusions. Further discussion of the limitations for data availability in South America or the understanding that “global” is not the most accurate descriptor for very large-scale conclusions improves the accuracy of these statements.

2. There are multiple conclusions in the paper that describe the presence of viruses in wastewater before they are reported in the population. While certain tobamoviruses were noted as being present around a year before their official report, EV-D68 was only reported during 2018- the same time as the spread of disease. If these viruses were mainly found in 2018, would that not shift the actual time frame of infection to early 2019 based on your previous conclusions? Not necessarily the primary focus of the paper but I think it is important to address, especially in the context of disease surveillance.

3. The phylogenetic trees are difficult to parse. Specifically Figure 8 with regards to astrovirus diversity, we don’t see the inclusion of the novel astrovirus BF34 as mentioned previously. If it is present in the phylogenetic tree, it is not obvious and may need to be reworked to be more consistent with the reported findings.

4. There are still red misspelling lines under the titles in figures S4 and S5.

5. Figure 5A is never actually referenced in the paper.

(Remarks on code availability)
Their code isn't anything special. There's nothing to review here.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have properly addressed my previous comments. Well done!

(Remarks on code availability)
Data and codes are available as pointed out by the authors in the European Nucleotide Archive.

Reviewer #3

(Remarks to the Author)
Myconcerns have been addressed

(Remarks on code availability)
Nothing to remarlk

We would like to thank the reviewers for their thorough review, and we have addressed the points raised by the reviewer point-by-point below.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The research led by Worp et al., analyzes DNA and RNA viromes from different parts (62 cities of the world. They use shotgun metagenomics combined with target enrichment viruses using GastroCap probe. The manuscript is overall well written. Wastewater surveillance mostly targets bacterial population. Studies conducted with viruses are less common mainly due incomplete or incorrect information on databases. I did find some gaps and major comments in their methodology. There is no prophage analysis. There are no clear distinction among RNA and DNA viruses in terms of discussion and results. I mean that makes difficult for the reader to point out patterns. Results are presented as "total/whole" viruses, maybe a different pattern could be observed if DNA and RNA viruses are presented separately. Especially when PCA are used to demonstrate variance across the different dataset here presented.

Thank you for your comments. We have performed additional analyses, and we hope that we have adequately addressed your comments.

There is no prophage analysis.
We address this in the comment about line 304 (see below).

There are no clear distinction among RNA and DNA viruses in terms of discussion and results. I mean that makes difficult for the reader to point out patterns. Results are presented as "total/whole" viruses, maybe a different pattern could be observed if DNA and RNA viruses are presented separately. Especially when PCA are used to demonstrate variance across the different dataset here presented.

We agree that distinguishing between RNA and DNA viruses could offer valuable additional insights, as they can differ in ecology, host range, and how they are processed during library preparation. To investigate this further, we performed additional principal component analyses (PCA) on the metagenomic dataset, analyzing DNA and RNA viruses separately (Supplemental Figure S6 and S7). These analyses were conducted at both the family and genus level and applied to biannual and longitudinal samples. However, after performing this analysis we found that the overarching conclusions presented in the manuscript remain robust. As reported, the original PCA on the total virome did not show clear clustering by continent but did reveal city-level variation. Tobamoviruses and Gemykibiviruses were the main contributors to variance in this analysis. Upon stratifying the data, these taxa remained among the dominant drivers in their respective genome-type-specific analyses (e.g., Tobamoviruses in the RNA-only PCA). Stratifying by genome type provided modest additional resolution. In the DNA-only analyses, city- and continent-level clustering became slightly more apparent in both longitudinal and biannual datasets. The RNA-only analyses showed more overlap between cities and offered further insight into the RNA viral taxa that contributed most to the observed variance. These stratified analyses offer complementary resolution but do not alter the main findings. However, we also noted that splitting the data introduces certain limitations. Since most viral families are either DNA or RNA viruses, it becomes more difficult to evaluate global

patterns in an integrated way. Our total virome PCA allowed us to assess compositional variance across locations based on the full virome, which aligns with our main objective to uncover broad spatial trends. Therfore, we decided to keep Figure 4 as it is, while including the new analysis and figures as Supplementary Figure S6 and S7.

In addition, to support interpretation of genome-type contributions, we updated Figure 2 (heatmap of viral families) and its caption to indicate whether each viral family consists of DNA or RNA viruses, thereby clarifying the genomic composition in both the metagenomic and capture datasets.

To make this clear in the manuscript, we have added the following to the Results section (line 156-158):
"This pattern was consistent when repeating the PCA separately for DNA and RNA viruses (Supplementary Figures S6 and S7), supporting the robustness of these findings across viral genome types"

In addition, we have added the following to the method section (line 480-481):
"PCA was also performed separately for DNA and RNA viruses to explore potential differences in community composition between viral genome types."

Line 113: Is this value normal for viruses. What were the remaining reads classified as? Bacteria, human reads, microeukaryotes?. I had to see about the distribution in one of the figures.

The observed percentages of viral reads in our dataset are consistent with previous studies of metagenomic sequencing of urban wastewater. In shotgun sequencing without enrichment, viral reads typically comprise a small fraction of total reads (often <1% to ~25)[1,2]. Capture-based approaches such as VirCapSeq VERT or the TWIST Comprehensive Virus Research Panel have been shown to substantially enrich viral content, with increases in viral read counts ranging from several-fold to over a thousand-fold in wastewater samples.[1,3,4] The remaining reads in our dataset were classified as bacterial, human or other eukaryotic, or remained unclassified (i.e., without any annotation). We have revised the text at line 116 for clarity and now explicitly refer to Supplementary Figure S2, which shows the composition of the reads at the superkingdom level. Furthermore, we now report the interquartile range (IQR) instead of the minimum and maximum, as it more accurately represents the central distribution of the data. This adjustment has also been applied to the reporting of read numbers in the manuscript.

Revised text (line 110-116):
The median number of reads generated after sequencing was 9.8×10^5 (IQR: 6.9×10^5-1.26×10^6) per sample for the metagenomic sequence dataset and 7.4×10^5 (IQR: 4.6×10^5- 1.17×10^6) for the capture sequence dataset (Supplementary Figure S2). The median percentage of reads per sample that could be annotated as viral was 11% (IQR: 6.5-15.3%) for the metagenomic sequence dataset and 26% (IQR: 11.8-46.9%) for the capture sequence dataset. The remaining reads were annotated as bacterial, human, other eukaryotic, or unclassified (i.e., without any annotation) (Supplementary Figure S2).

Lines 171-172: Australia compared to other continents is not interconnected to any other one mass of land. Something that may be discussed/highlighted in the manuscript.

We agree that Melbourne's geographic positioning may play a role in shaping its distinct virome composition, potentially influencing patterns of viral introduction and circulation. We have added a sentence to acknowledge this potential contributing factor in the revised manuscript.

Revised text (Line 173-175):
At the genus level, the enteric virome composition in Melbourne (Australia) was distinct from that of all other cities, primarily due to a high abundance of Mamastroviruses which may reflect differences in interregional connectivity (Figure 5D).

Line 304: Was the viral DNA fraction (aka phages) in this study screened for ARGs or prophages. That may provide valuable complementary information.

We thank the reviewer for this suggestion and acknowledge that bacteriophages can contribute to antimicrobial resistance (AMR) via generalized transduction or by integrating into bacterial genomes as prophages. To address the reviewer's point, we screened all assembled viral contigs (>300 bp) from both shotgun and capture-based datasets for integrated viral elements (i.e., prophages) and for antimicrobial resistance genes (ARGs) using geNomad[5]. We did not detect any prophages. We note that our sample preparation protocol, including filtration and treatment with Omnicleave, was designed to enrich for viruses, thereby reducing the amount of bacterial host DNA that could contain prophages. This may, in part, explain why no prophages were detected in our dataset.
Only three Caudoviricetes phage contigs showed ARG-related annotations:

- Two contigs (28 kb and 18 kb) from samples from Seattle (shotgun metagenomics and capture) contained a 23S rRNA methyltransferase gene (cfr), associated with linezolid resistance.
- One contig (3.8 kb) from a Chapel Hill sample (shotgun metagenomics) encoded a rifampin ADP-ribosyltransferase, associated with rifampin resistance.

These findings were limited to a small number of contigs (3) and samples (2).
As *Caudoviricetes* include both lytic and temperate phages, the presence of ARGs warrants careful interpretation. These findings are in line with the limited evidence for phage-encoded ARGs[6] and support the cautious interpretation of such annotations. We note that the main scope of our study was to characterize viral diversity in urban wastewater using metagenomic approaches, rather than to investigate phage-mediated AMR specifically. Nonetheless, we appreciate the opportunity to clarify this point.

We have added the details of the prophage and AMR analysis to the Method section line 456-458:
"Additionally, viral contigs ≥300 bp were screened for antimicrobial resistance genes (ARGs) and prophages using geNomad (version 1.11.0)[5] with default settings. No prophages were detected, and only three contigs contained ARG annotations (Supplementary table 1)."

We also acknowledged these results in the discussion line 306-312:
"Additionally, an expanded dataset of these wastewater samples -sequenced using a strategy optimized for bacterial and AMR profiling- showed distinct regional variations in the resistome and bacteriome across continents. Specifically, bacterial composition divides the world's regions into roughly two major groups (Europe, Central Asia & North America versus Africa & Middle East), while

the resistomes showed more unique regional profiles. *Our analysis revealed very limited ARG presence in the viral fraction, which may be due to the virus-enrichment protocol and the infrequent occurrence of ARGs by bacteriophage genomes*[6] *(Supplementary Table 1).”*

Line 340: Figure? or point out the months. Any particular parameter that could have driven this pattern.

Line 340 refers to Figure 7B. We have added a reference to this figure to clarify the observed patterns. While seasonal peaks in HAstV infections have been reported — typically during colder months in temperate regions such as China, Spain, and the USA[7–9]. or during the rainy season in tropical regions[10] these studies generally report total HAstV prevalence and do not resolve trends by genotype. A comprehensive understanding of genotype-specific temporal dynamics remains limited. The type-specific differences observed in our study could indicate type-specific seasonal trends. However, identifying specific drivers would require integration of temporally resolved environmental data (e.g. temperature, humidity), host population characteristics (e.g. age structure, immunity), and population-level factors such as density or mobility — data that are currently unavailable or limited for most sampling sites.

We have revised the text:

Revised text (line 343-348):
“Our findings showed consistent detection of some astroviruses like HAstV-1 in most locations with longitudinal sampling, while other types such as MLB1, were only detected in specific months (Figure 7B). These findings suggest that astrovirus seasonality may be type specific. The type specific differences observed in our study may reflect a complex interplay of viral characteristics (e.g. shedding duration, tropism), host population characteristics (e.g, population immunity, age distribution), and potentially environmental factors (e.g. temperature, humidity, rainfall).[7–10]”

Line 392: Value should be expressed in gravities instead of rpm.
We adjusted the text in the manuscript in line 402 to replace 4000 rpm with 3744×g and added extra details on the concentration step in line 434-437 .

Revised text (line 402-408):
“Samples were thawed at room temperature and 50 ml of raw wastewater was centrifuged at 3744 ×g for 15 min to remove large debris. The supernatant was filtered using 0,45 µm pore diameter filter to remove bacterial and eukaryotic cells. Although 0.22 µm filters are also used in virome studies, they have been associated with reduced recovery of certain viral families and were therefore not used in this study.[11] A volume of 30 mL of filtered supernatant was concentrated using 30 kDa Pierce™ Protein Concentrators PES (Thermo Fisher Scientific) by sequential centrifugation at 6000 ×g at 4 °C (15 minutes for the first 15 mL, followed by 30 minutes for the second 15 mL added to the same concentrator).”

Line 393: In addition 0.2 um filter could have been used. It is usual to get 16S rRNA gene or material still passs thru 0.45 um filters. While I acknowledge the use of OmniCleave there is no percentage of recovery or the efficiency of this commercial product to remove free DNA that could pose as contamination in the "viral" fraction. In addition, I did not see any experiment regarding a

mock viral community that could have helped to present recovery rate plus sequencing control.

Filtration through 0.22 µm filters are commonly used to reduce bacterial and host DNA contamination, and are commonly used in virome studies. However, studies have reported a reduction of the viral fraction when using filters with this size. Matthijnssens et al used mock viral and bacterial communities and showed that although both 0.22 µm and 0.45 µm filters efficiently removed bacterial DNA, 0.22 µm filters resulted in significantly higher losses(>90–99%) of several viral families such as Herpesviridae and Mimiviridae, as well as smaller viruses including Rotavirus, Parvovirus, and Circovirus. Notably, the difference in 16S rRNA reduction between the two filter sizes was modest. These results support the concern that 0.22 µm filtration offers limited added benefit in reducing bacterial 16S rRNA signal, while substantially compromising viral recovery. Therefore, we selected a 0.45 µm PES filter, which is also commonly used in virome studies[11–13] and complemented this with OmniCleave nuclease treatment to further reduce free extracellular bacterial and host DNA/RNA.

To make this clear we have added the following sentence to the Methods section (line 403-405):
"The supernatant was filtered using 0.45 µm pore diameter filter to remove bacterial and eukaryotic cells. Although 0.22 µm filters are also used in virome studies, they have been associated with reduced recovery of certain viral families, and were therefore not used in this study.[11]"

To assess the effect of OmniCleave, we performed replicate sequencing of the same sample with and without treatment and visualized the relative superkingdom-level read composition in both the shotgun metagenomics and capture datasets (Supplementary Figure S9). The comparison shows a clear reduction in the fraction of bacterial and eukaryotic (including human) DNA reads and an increase of the viral fraction. These findings support the treatment's efficacy in reducing non-viral genetic background align with previous studies demonstrating that enzymatic treatments can enrich viral signal by depleting host and bacterial nucleic acids in metagenomic samples.[13]

While mock viral communities can be informative for benchmarking extraction and sequencing protocols, and spike-in controls such as mengovirus are commonly used in (RT)-qPCR workflows to assess recovery and inhibition, their role in untargeted metagenomic studies is more limited. In complex matrices like wastewater, achieving a consistently low and quantifiable level of spiking is technically challenging due to the difficulty of maintaining detectable spike-in levels without distorting the sample. However, we agree that benchmarking with internal controls is valuable for method development. Therefore, to monitor potential contamination and assess reproducibility, we included both negative and positive controls in our workflow. Negative controls (nuclease-free water) were processed alongside samples, and one sample was sequenced multiple times as a technical replicate. As shown now in the new Supplementary Figure S9, the metagenomic dataset showed comparable superkingdom-level composition across replicates, despite variability in read numbers. Also in the capture dataset, samples with similar read depth showed comparable composition, supporting the consistency of the approach.

To clarify this we have added a paragraph to the Methods section (line 441-443):
"To monitor potential contamination and technical consistency, we included negative controls and resequenced one sample as a technical replicate for each batch. As shown in Supplementary Figure S9, replicates with similar read depth exhibited comparable taxonomic composition."

Reviewer #1 (Remarks on code availability):

The code is presented, but no data are currently not available. Thus, I could not tried the code presented by the authors.

The data was made available on the European Nucleotide Archive under project: PRJEB87273.

Reviewer #2 (Remarks to the Author):

The study describes the detection of viral species prevalent in wastewater obtained from 62 cities located in six different continents. The study primarily used two approaches: shotgun metagenomic approach and a targeted approach to capture enteric viruses. The study revealed the vast diversity of viruses that infect a wide range of hosts spanning all major kingdoms. As expected, the capture approach enriched the targeted viral families. Interestingly, the presence of certain groups of viruses were temporally correlated in some cases and in others clearly associated with a geographical location. This was further corroborated using PCA analysis revealing unique signatures associated with locations (cities) and the time of collection (seasonal). The study identified distinct viral diversity patterns and abundant viruses that influenced viral composition, which is relevant in plant, animal, and human health. The detection of plant viruses, especially those that appear to be emerging viruses, is important for agriculture and food security. Viruses detected only in wastewater indicated the inability of clinical surveillance in capturing the full diversity of human viruses. Wastewater monitoring also allowed tracking of enteric viruses such as polio, which is of immense clinical importance. The study also highlighted the presence of a large proportion of unassigned viral taxa, which can lead to the discovery of novel or emerging viruses. This study is an important step in understanding the application of wastewater-based monitoring in informing human, animal and even plant health.

Thank you for your comments.

Reviewer #3 (Remarks to the Author):

This paper sought to understand the potential for wastewater analysis as a method of disease and virus surveillance through a combination of shotgun metagenomic sequencing and GastroCap sequence probes. This approach addresses the need for better disease surveillance within cities and utilizes an approach that can be both cost effective and easy to implement. Samples were collected from a variety of geographical locations across multiple seasonal time frames, highlighting the consistency of the global viral community within wastewater environments. Additionally, shorter time frame sampling was able to provide insight into the more fine-tuned viral profile within cities, showing distinct differences between the families of virus present in different areas. The diversity of reported viral families and genera covers many well-defined groups in addition to several potentially novel viruses, making this of particular interest to virologists and molecular epidemiologists. The potential for this type of surveillance to improve public health is significant, as several of the collected samples across the globe were able to accurately predict specific outbreaks multiple months before they were first reported.

Comments:

The viral metagenomics analysis performed is very thorough and they very rigorously characterize viral diversity.

Thank you for your comments. We have addressed your comments point-by-point below:

MinorNotes:

1. While there is an amount of variety in the geography of these samples, I would argue that the lack of data for South and Central America impacts the meaning of your "global" and South American-centric conclusions. Further discussion of the limitations for data availability in South America or the understanding that "global" is not the most accurate descriptor for very large-scale conclusions improves the accuracy of these statements.

Our study aimed to characterize the urban wastewater virome using a dataset with broad geographic coverage, spanning six continents and 62 cities representing diverse urban contexts. We use the term "global" to reflect this broad geographic scope and the aim of exploring large-scale patterns in virome composition. However, we acknowledge that the dataset is not completely geographically exhaustive. Sample collection was based on voluntary contributions to the Global Sewage Surveillance network and shaped by the availability of local partners and infrastructure needed to collect, store, and share samples. This resulted in some geographic imbalances, including underrepresentation of certain regions—particularly South and Central America in this study. To address this, we included a statement in line 304 the Discussion clarifying that "sampling in South America, Central America and Oceania was limited, making the findings less generalizable for these regions." We also reviewed and refined wording throughout the manuscript to ensure that the scope of our conclusions accurately reflects the dataset.

2. There are multiple conclusions in the paper that describe the presence of viruses in wastewater before they are reported in the population. While certain tobamoviruses were noted as being present around a year before their official report, EV-D68 was only reported during 2018- the same time as the spread of disease. If these viruses were mainly found in 2018, would that not shift the actual time frame of infection to early 2019 based on your previous conclusions? Not necessarily the primary focus of the paper but I think it is important to address, especially in the context of disease surveillance.

We appreciate the reviewer's observation. In the case of tobamoviruses, we detected Tomato brown rugose fruit virus (ToBRFV) in wastewater samples collected prior to its first published reports, illustrating the potential of wastewater metagenomics to flag emerging plant pathogens before they are captured through plant health surveillance programmes or scientific reports. Similarly, for Echovirus 30 (E-30), detection in European wastewater samples preceded clinical outbreak reports, suggesting earlier circulation. However, Enterovirus D68 (EV-D68) was primarily detected in 2018, consistent with its known biennial circulation pattern in temperate regions. In this case, wastewater detections aligned with, rather than anticipated, clinical observations. These findings support the utility of wastewater metagenomics in capturing circulating viruses, but indeed do not imply early detection for every pathogen. The early-warning potential of wastewater metagenomics depends on

virus-specific factors such as shedding dynamics, stability in wastewater, transmission route, and also the extent of existing surveillance.

We have added a section to the discussion in lines 358–362 to reflect this point:
"Unlike some other viruses, wastewater detections of EV-D68 coincided with, rather than preceded, the timing of clinically recognized circulation. These findings suggest that the potential of wastewater metagenomics for early detection varies by virus and is influenced by biological factors, such as shedding dynamics, stability, and transmission mode, as well as clinical surveillance-related factors."

3. The phylogenetic trees are difficult to parse. Specifically Figure 8 with regards to astrovirus diversity, we don't see the inclusion of the novel astrovirus BF34 as mentioned previously. If it is present in the phylogenetic tree, it is not obvious and may need to be reworked to be more consistent with the reported findings.

We will share the high resolution trees directly with the editor, as they became lower resolution during submission. Figure 8 is a phylogenetic tree of partial ORF2 of Human Astrovirus type 5 only, and BF34 is indeed not included. Figure S8 is comprised of partial ORF2 sequences of Human Astrovirus type 1-8 and also does not include BF34. The sequences available for BF34 and other novel astroviruses (e.g. MLB1) were sparse and corresponded to different, non-overlapping regions of ORF2, and could therefore not be included in a single phylogenetic analysis aimed at comparing global astrovirus diversity across types.

4. There are still red misspelling lines under the titles in figures S4 and S5.

Thank you for pointing this out. We adjusted this in figures S4 and S5.

5. Figure 5A is never actually referenced in the paper.
We have added a reference to figure 5A, showing the PCA of enteric viruses at the family level for the biannual samples to line 159-161.

Revised text (line 159-161):
"Focusing on the enteric virus families in the GastroCap sequence dataset, PCA revealed a broadly composition of these viral families across all continents, as indicated by overlapping clusters in the plot (Figure 5A)."

Reviewer #3 (Remarks on code availability):

Their code isn't anything special. There's nothing to review here.

References

1. Tisza, M. et al. Wastewater sequencing reveals community and variant dynamics of the collective human virome. Nature Communications 2023 14:1 14, 1–10 (2023).

2. Nieuwenhuijse, D. F. et al. Setting a baseline for global urban virome surveillance in sewage. Scientific Reports 2020 10:1 10, 1–13 (2020).

3. Schaeffer, J., Desdouits, M., Besnard, A. & Le Guyader, F. S. Looking into sewage: how far can metagenomics help to detect human enteric viruses? Front Microbiol 14, (2023).

4. Martínez-Puchol, S. et al. Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens. Science of The Total Environment 713, 136604 (2020).

5. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. Nature Biotechnology 2023 42:8 42, 1303–1312 (2023).

6. Enault, F. et al. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. The ISME Journal 2017 11:1 11, 237–247 (2016).

7. Luo, X., Deng, J. kai, Mu, X. ping, Yu, N. & Che, X. Detection and characterization of human astrovirus and sapovirus in outpatients with acute gastroenteritis in Guangzhou, China. BMC Gastroenterol 21, (2021).

8. Guix, S. et al. Molecular epidemiology of astrovirus infection in Barcelona, Spain. J Clin Microbiol 40, 133–139 (2002).

9. Chhabra, P. et al. Etiology of Viral Gastroenteritis in Children <5 Years of Age in the United States, 2008–2009. J Infect Dis 208, 790–800 (2013).

10. Bosch, A., Pintó, R. M. & Guix, S. Human astroviruses. Clin Microbiol Rev 27, 1048–1074 (2014).

11. Conceição-Neto, N. et al. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. Scientific Reports 2015 5:1 5, 1–14 (2015).

12. Fernandez-Cassi, X. et al. Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. Sci Total Environ 618, 870–880 (2018).

13. Hall, R. J. et al. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. J Virol Methods 195, 194–204 (2014).