Water Research 285 (2025) 123985

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Comparative wastewater virome analysis with different enrichment methods

Matthew Thornton [a], Gabriela Eder [b], Fabian Amman [a,c], Anastasiia Pantielieieva [a], Julia Vierheilig [b,d], Andreas Bergthaler [a,c,*]

[a] Department of Hygiene and Applied Immunology, Medical University of Vienna, Kinderspitalgasse 15, Vienna 1090, Austria
[b] Institute of Water Quality and Resource Management, TU Wien, Lilienthalgasse 21 OD, Vienna 1030, Austria
[c] Research Center for Molecular Medicine of the Austrian Academy of Science, Lazarettgasse 14, AKH BT 25.3, Vienna 1090, Austria
[d] Interuniversity Cooperation Centre Water & Health, Austria

ARTICLE INFO

Keywords:
Wastewater surveillance
Wastewater-based epidemiology
Virus
Enrichment
Concentration
PCR
Sequencing
Virome

ABSTRACT

Wastewater-based epidemiology (WBE) has proven its value for public health. Physical concentration of virus particles is a crucial step for WBE to permit a sensitive and unbiased characterization of the catchment virome. Here we evaluate five different virion concentration techniques, including polyethylene glycol precipitation (PEG), vacuum-based direct capture (VDC), ultrafiltration (UF), NanoTrap® (NT), and membrane adsorption (MEM) for their suitability to concentrate a wide variety of viral taxa from raw wastewater for PCR detection and sequencing-based metagenomic readouts. We found that to capture a taxonomically diverse virome from wastewater, PEG and VDC outperform all other methods tested in enrichment rates, reproducibility, species detection, and captured nucleotide diversity. We observed that different methods exhibit variable concentration efficiencies across taxonomic groups in a reproducible manner, though we could not identify common physiochemical attributes driving this difference. We conclude that both PEG and VDC are equally capable at detecting and enriching a broad range of viral taxa, boosting the genomic information potential and reducing blind spots relative to other tested methods. These results advance WBE towards capturing the complex wastewater virome and help guide protocol choices for potential future viral threats.

1. Introduction

Wastewater (WW) is increasingly recognized as a valuable source of information about the health of the contributing population (Choi et al., 2018) including drug use (Huizer et al. 2021), chemical exposure (Vitale, Morales Suárez-Varela, and Picó 2021), biomarkers (Gracia-Lor et al. 2017), and pathogen surveillance (Polo et al. 2020; Kilaru et al. 2023). Of particular interest are viral pathogens, which are shed into the wastewater catchment through feces, urine, and other bodily fluids, and transported to centralized sampling points, typically at local wastewater treatment plants (WWTPs). Such samples can provide an unbiased view of the viruses circulating in a population. Notably, this is not hindered by constraints such as available access and participation in a healthcare system, and symptomatic disease manifestation present in clinical sampling strategies. Wastewater sampling overcomes these limitations by passively sampling the population, assuming the population is connected to centralized wastewater infrastructure. Multiple studies have shown high concurrence between wastewater load and clinical load (Melvin et al. 2021; Amman et al. 2022; Ahmed et al. 2023; Tisza et al. 2023). This has been successfully applied to monitor and track the spread of SARS-CoV-2 variants and outbreaks (Amman et al. 2022; Karthikeyan et al., 2022; Jahn et al., 2022), Mpox (Oghuan et al. 2023; Sherchan et al. 2023), Influenza A (Wolfe et al. 2022), Poliovirus (Tambini et al. 1993), and many other pathogenic targets (Kilaru et al. 2023).

To address the highly diluted nature of viruses in raw wastewater, concentration methods to enrich virus particles are usually applied before nucleic acid extraction. Ideally, a suitable concentration method for a representative surveillance of the circulating virome specifically enriches virus particles regardless of physicochemical characteristics (e.g. size, envelope status, isoelectric point, nucleic acid type, shape) while simultaneously reducing the background of non-viral material and inhibitors. Various approaches have been proposed and are routinely used in wastewater pathogen surveillance programs worldwide (Carcereny

* Corresponding author.
E-mail address: andreas.bergthaler@meduniwien.ac.at (A. Bergthaler).

https://doi.org/10.1016/j.watres.2025.123985
Received 6 March 2025; Received in revised form 22 May 2025; Accepted 7 June 2025
Available online 8 June 2025
0043-1354/

et al., 2021; Amman et al. 2022; Adams et al. 2024; Krogsgaard et al., 2024). These methods differ in their mechanism, viral specificity, ease of use, and scalability. A growing body of literature aims to evaluate the robustness and efficiency of these methods for concentrating virions from wastewater. These studies focus largely on respiratory virus targets like SARS-CoV-2, or surrogates thereof, typically using quantitative PCR as a readout. Previous work comparing viral concentration methods has shown that polyethylene glycol precipitation (Hjelmsø et al. 2017; Barril et al. 2021; Farkas et al. 2021; Salvo et al. 2021), ultrafiltration (Jafferali et al. 2021; Pino et al. 2021; Farkas et al. 2022; Zheng et al. 2022), and adsorption-elution with an electronegative membrane (Ahmed et al. 2020; LaTurner et al. 2021) yield high recoveries of the measured viral targets. More recent methods such as NanoTrap® and vacuum-based direct capture, based on Promega's VacMan®, were also shown to perform well for certain viral targets (Dimitrakopoulos et al. 2022; Girón-Guzmán et al. 2023; Antkiewicz et al. 2024). Given the variable physicochemical characteristics of virus particles across the viral kingdom, the generalization of these results to newly considered viruses are in question. To date few studies have investigated the effects of concentration methods on the total virome using sequencing based read outs (Hjelmsø et al. 2017; Jiang et al. 2024; Li et al. 2024; Cheshomi et al., 2024). These methods suffer from more delicate molecular biological protocols and are more impacted by inhibitor presence and sample composition. At the same time, metagenomic sequencing allows capturing of the full viral diversity, and is already used in WBE applications to track a wide variety of viruses. Assessing the suitability of different concentration methods across this wide range of taxa not only allows evaluation of their suitability for metagenomic applications but can also inform method choices for future emerging viral pathogens.

To evaluate the process of enriching viral nucleic acids at the virome scale, we set out to comparatively analyze five different concentration methods using orthogonal readouts of dPCR against 11 morphologically diverse viral targets, combined with deep sequencing to evaluate overall viromic data quality and identify method-associated biases across a broad spectrum of viral taxa. We then compared the two best in-class methods to confirm the reproducibility of the observed signals across different WWTPs of varying contributing population sizes and different wastewater matrices, offering valuable generalizable insights for future viral WBE applications.

2. Materials and methods

2.1. Wastewater collection

Samples were collected from the inflows of three different municipal WWTPs within Central Europe using activated sludge treatment including nitrification, denitrification and phosphorous removal. Sampling was done volume-proportionally (c.v.v.t – constant volume variable time) using an automatic sampler over a period of 24 h. The composite samples were subsequently transported to the laboratory where samples were stored at 4 °C until further use. All samples were processed within 48 h.

2.2. Experimental setup

To evaluate the performance of different virus particle concentration methods across the community of viruses in wastewater, we started out with a 24-hour composite wastewater sample from a large, municipal WWTP (with a design capacity of >2 million population equivalents), treating the combined sewer effluent of a big city. Supplemental Figure S1 visualizes the experimental setup.

A viral spike-in mix was created consisting of five viruses, selected to include enveloped/non-enveloped, different sizes, genetic nucleic acid types, and shapes by mixing individual stocks of Influenza A (PR-8 strain), Vaccinia (Western reserve strain), LCMV (Clone 13 strain), LCMV (Armstrong strain), Bacteriophage MS2, and PhiX174 at a target concentration of 1×10^5 genome copies/mL. For details, refer to Table S1. The spike-in mix was added to 600 mL of raw wastewater and equilibrated for 2 h at 4 °C with regular inversion every 10 min to ensure adequate mixing. An equal volume of spike-in mix was added directly into a 2x volume of DNA/RNA shield for direct TNA extraction. Spike-in quantifications per target were between 2.8×10^3 and 1.4×10^4 genome copies per milliliter (Suppl. Table S3).

The spiked wastewater was subjected to five different concentration methods: Membrane adsorption (MEM), ultrafiltration (UF), NanoTrap® (NT), vacuum-based direct capture (VDC), and polyethylene glycol precipitation (PEG). Negative controls of an equal volume of H_2O (NTC) were run through each concentration protocol to control for lab contaminants and a non-spiked WW sample (NSC) served as a control for background levels of the spike-in viruses already present in the WW. The concentrated wastewater was then used as input for total nucleic acid (TNA) extraction along with three replicates of 500 μL of raw, unconcentrated (UNC) wastewater from both the spiked and non-spiked WW samples and an extraction control (H_2O). The unconcentrated extracts served as an unprocessed reference, which was expected to be less sensitive due to the reduced input volume, but unaffected by method-specific bias, or sample loss associated with the additional handling steps.

Subsequently, all samples were analyzed for selected viral targets by dPCR and for the bacterial 16S rRNA gene by qPCR (see below) and prepared for sequencing both by untargeted metagenomics and hybrid-capture deep sequencing using the Twist Comprehensive Viral Research Panel (see below).

Finally, we tested the robustness of our findings by expanding our analyses beyond samples of a single large, urban WWTP by including wastewater samples from a second time point and two additional WWTPs in our analysis. To cover different wastewater matrices, we selected a medium-sized, semi-urban WWTP and a small, rural WWTP, sewer catchment, with a design capacity of 55,000 and 10,000 population equivalents respectively, all located within the same geographic region in Central Europe. Samples from these three WWTPs were concentrated with PEG and VDC in triplicates for a total of 18 samples and analyzed by hybrid-capture sequencing.

2.3. Concentration methods

From the spiked wastewater stock, triplicates of 40 mL were aliquoted per concentration method, though for NT only 10 mL was aliquoted following suggestions by the manufacturer. After an initial step for pelleting down large solids from the WW, ultrafiltration (UF) was carried out using Amicon Ultra-15 (30 kDa) Centrifuge Filter tubes to concentrate 40 mL to a volume <0.5 mL which was directly put into the TNA extraction. For the NanoTrap® (NT) concentration method, virus particles were enriched with Ceres Nanotrap® Microbiome A magnetic particles according to the manual from the manufacturer. MEM samples were processed following a vacuum-filtration protocol with 0.45 μm mixed cellulose ester membrane (Merck Millipore, HAWP04700) as described in Ahmed et al. (2020) with slight modifications. PEG precipitation was conducted as described in Radu et al. (2022) with a few adaptations. Silica membrane adsorption based viral enrichment using a vacuum-based direct capture device (VacMan®, Promega) was conducted according to manufacturer's protocol with a few adjustments. Details on the different concentration methods are given in the supplemental text.

2.4. TNA extraction

TNA extraction was conducted using a magnetic bead based, automatic extraction device (Maxwell RSC; Promega) with the "RSC PureFood GMO and Authentication Kit" (Promega). The cartridges were prepared according to the manufacturer's instructions. For unconcentrated samples as well as samples enriched using PEG, VDC, UF, and

MEM, 500 μL lysis buffer were added to the extraction cartridges. For NT samples, no additional lysis buffer was added. Samples were eluted in 100 μL elution buffer. PEG and VDC samples from the repetition experiment were eluted in 60 μL.

2.5. PCR assays

Quantification of five spiked and six endogenous viruses was performed by dPCR. SARS-CoV-2 and PMMoV (as well as the internal amplification control, IAC) were measured on the Applied Biosystems QuantStudio Absolute Q Digital PCR System (Thermo Scientific) with 20.5k partition plates while all other targets were measured on the QIAcuity dPCR system (QIAGEN) with 26k partition plates. Inhibition of the dPCR reactions was assessed by comparing the IAC concentrations in the extraction control (H_2O) to those in the samples, which didnt show considerable effect on viral quantifications (Suppl. Fig. S2D). Bacterial 16S rRNA gene (V3-V4 region) quantification was conducted by qPCR on the Applied Biosystems™ QuantStudio™ 6 Pro Real-Time PCR (Thermo Fisher Scientific). To address potential inhibition effects in the qPCR, we applied a dilution approach (for details refer to the supplemental text). Details on the PCR targets and their viral characteristics can be found in Suppl. Table S1 and information on cycling conditions, primer and probe information and on analysis in the supplemental text as well as in Suppl. Table S2. Background levels of each spike-in virus were also measured from the non-spike control WW samples. Measurable levels of background signal were negligible and never surpassed 2 % of the total spiked signal (Suppl. Fig. S2). For this reason, background levels were not considered as a source of bias in downstream dPCR analysis.

2.6. Library preparation and hybrid-capture

15 μL of each extract was used for input to the library preparation. Reverse transcription was carried out with New England Biolabs (NEB) Protoscript II (NEB, E6560L) and NEB Second Strand Synthesis Module (NEB, E6111L) following manufacturer's protocol. Each cDNA product was equalized to 100 ng as input to the NEBNext Ultra II DNA FS Library Prep (NEB, E7805L) protocol and subsequent hybrid-capture with the Twist Comprehensive Viral Research Panel (Twist Bioscience). Full details can be found in the supplemental text. Both hybrid-captured and untargeted shotgun libraries were sequenced with a paired-end, 200 cycle NovaSeq SP flow cell at the Biomedical Sequencing Facility of the Research Center for Molecular Medicine of the Austrian Academy of Sciences (CeMM).

2.7. Data and code availability

Sequencing results were processed with an in-house pipeline adapted from Tisza et al., 2023. Full details concerning the workflow and subsequent analyses are found in the supplemental text. All code used in the analysis and figure production can be viewed on Github (https://github.com/thorntonmr/comparative_ww_virome_analysis_manuscript). The data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB86516 (https://www.ebi.ac.uk/ena/browser/view/PRJEB86516).

3. Results

3.1. Quantification of selected viruses and bacterial background by dPCR/qPCR

A comparison of the viral quantifications measured by dPCR for six endogenous and five spiked viruses highlights the pronounced differences between extracts from different virus concentration methods (Fig. 1A, Suppl. Fig. S2). Concentration performance was assessed for each target and defined as the ratio of target concentration per concentrated sample to a theoretical, expected concentration deduced from the measured concentration in the unconcentrated samples and the change in volume for each method from input to extraction (see equation 1 in the supplemental text). As the unconcentrated values should be free from concentration method-specific biases this provides a comparative reference value between methods and allows us to expand our comparison across both spike-in and endogenous viral targets. Thereby the metric indicates the fraction of virus which is preserved during the concentration procedure compared to a workflow without any concentration step. Calculating this expected value requires the detection of the target in the unconcentrated extract, which can be limiting for low abundance targets, but in our data only rotavirus A was below the limit of detection in the UNC extracts (Suppl. Fig. S2A).

VDC had the highest mean concentration performance followed by NT and PEG (0.91, 0.64 and 0.48 respectively; Fig. 1A). Performances varied between methods with VDC, PEG, and NT outperforming UF and MEM for most targets except for Vaccinia, where NT and MEM showed better performance compared to the other enrichment methods (Fig. 1A, Suppl. Fig. S2, Suppl. Table S1). As targets covered a large range of concentrations spanning over five orders of magnitude, we also investigated the effect of target concentration on performance. We found that performance was largely unaffected by concentration (Suppl. Fig. S3A). When comparing reproducibility between technical replicates, all methods produced comparable coefficients of variation with PEG exhibiting the most uniform variation across all the viruses (Suppl. Fig.

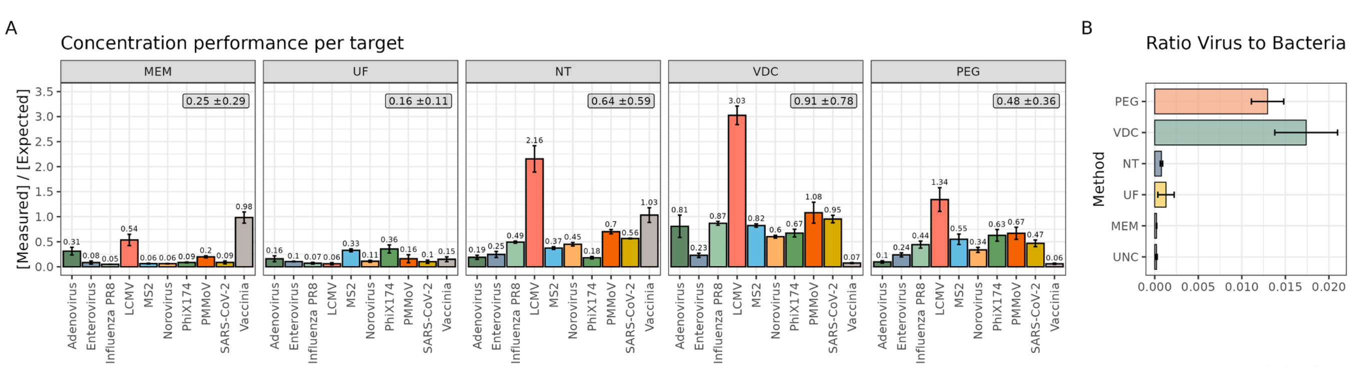

Fig. 1. Quantification of target viruses by dPCR across different concentration methods. Analysis of target virus quantification from wastewater extracts by dPCR assays designed against spiked-in and endogenous viruses. (A) Concentration performance across ten target viruses in extracts from different concentration methods expressed as the measured concentration divided by the expected concentration ±standard deviation (error bars) across $n = 3$ replicates. Box information gives the overall average performance per method. (B) Display of the ratio between sum of all virus target quantifications and 16S rRNA gene content, as a proxy for concentration efficiency of respective methods for virus particles over bacterial cells.

S3B).

To estimate bacterial contamination, we used 16S rRNA gene quantifications (qPCR). PEG and VDC showed a reduced load of bacterial sequences relative to measured virus suggesting a higher specificity for concentrating viral particles (Fig. 1B).

3.2. Virome assessment by NGS

Hybrid-capture deep sequencing was used to investigate the impact of method selection on the virome detected in wastewater. The dPCR quantifications correlated with the relative abundance measures from the dPCR-tested viruses included in the hybrid capture panel indicating minimal bias introduction to the relative quantifications measured via sequencing ($R^2 = 0.47$–0.84 for different methods; Suppl. Fig. S5).

All methods had highly reproducibly species resolution data with PEG and VDC standing out in the number of species detected in the low abundance range (Fig. 2A). Analyzing the overall picture of sample compositions at superkingdom level revealed that all concentration methods increased the number of viral reads compared to the unconcentrated samples. PEG had the highest percentage of viral reads followed by VDC (28.9 %±3.73 and 23.8 %±0.83), as well as the lowest percentage of reads assigned to bacterial taxa (28.6 %±1.75 and 41 % ±1.3). UF, MEM, and NT all had roughly half the percentage of reads on target (12 %, 14.2 %, 15.4 %) and bacterial content load followed the trend observed in the 16S rRNA gene dPCR, suggesting the 16S rRNA gene quantification gives a reliable indicator of background sequence contamination. We can observe the importance of the concentration step here as on average only 2.8 % of all reads in the unconcentrated samples were classified as viral (Fig. 2B).

Expected species richness (SR) was calculated across the full dataset (all methods, all replicates) with Chao's lower bound estimator. We then compared the observed species count per method to this overall estimate to obtain the species richness ratio (SRR). PEG recovered 76 % of the maximum detected species number, followed by 67 % for VDC and 51 %

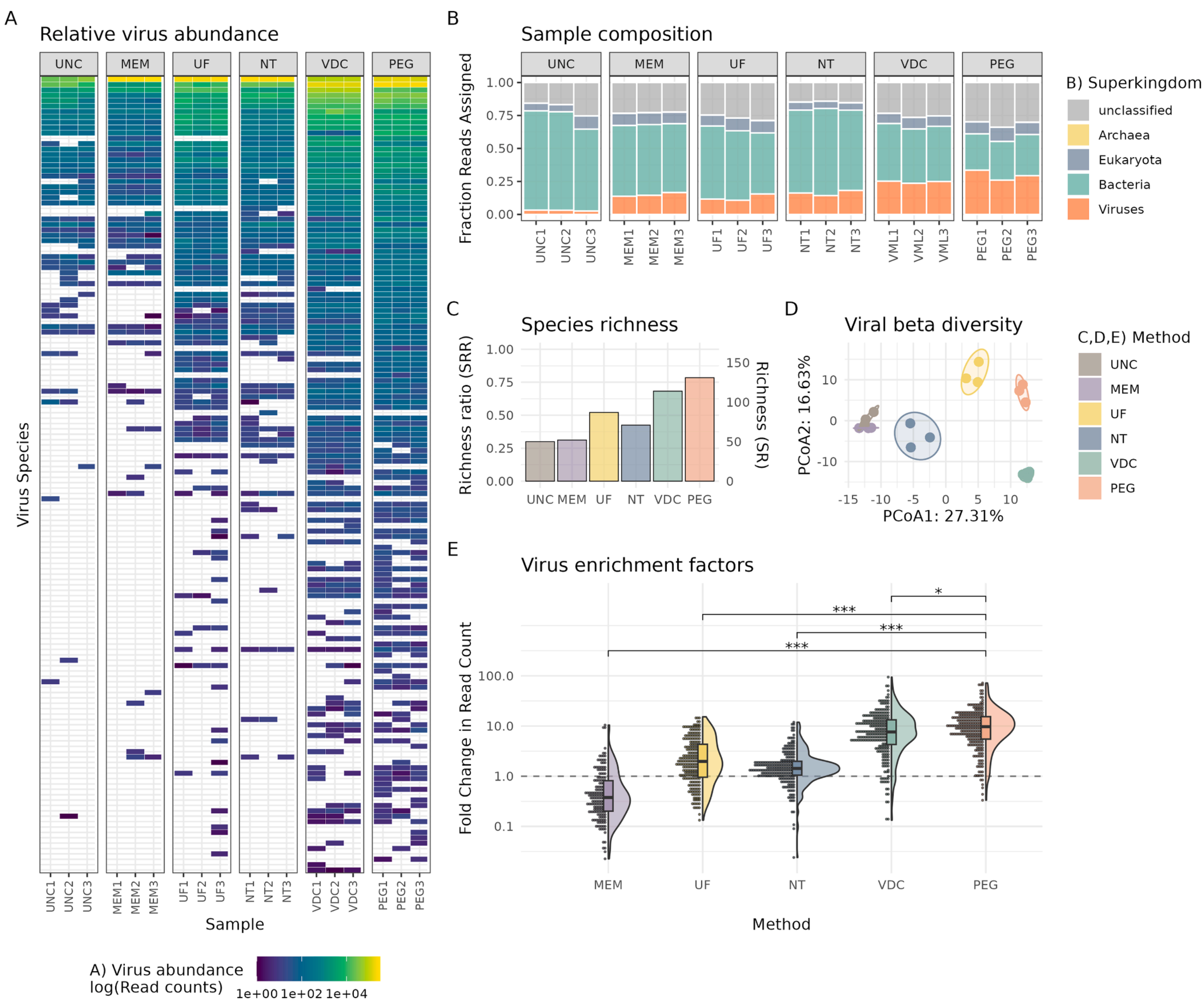

Fig. 2. Sequencing results of enriched viral libraries. (A) Heatmap showing viral species abundance with tile color indicating the log read count. Species names were removed for visualization purposes but are shown in Suppl. Fig. S6. White cells indicate that the species was not detected in that sample. (B) Composition of each sample by superkingdom via read assignment by Kraken2. (C) Species richness and richness ratio calculated using Chao's lower bound estimate. (D) Principal component analysis of the Aitchison distance for each method. Points are replicates of the same method. (E) Enrichment factor (read count ratio) of each method relative to the mean unconcentrated read count. Significance bars are only shown for comparisons to PEG for visualization purposes. Pairwise significance was tested with a Kruskal-Wallis rank-sum test and post hoc Dunn's test with the Holm-Bonferroni correction for multiple testing. Significance codes are as follows: * $p < 0.05$, *** $p < 0.001$.

for UF. All other methods detected fewer than 50 % of the theoretically detectable viruses (Fig. 2C). A virus was considered detected if our analysis considered it a reliable signal based on a minimal number of total unique reads and minimum genomic coverage (see supplemental text Bioinformatic analysis for details).

Given that the concentration methods differ in their mechanism of action, the resulting viral community recovered by each method may differ in both composition and abundances of its individual taxa. To assess this, beta diversity measures were calculated for the viral communities of each method and visualized with a principal coordinate plot, showing that the replicates per method separate in distinct, tight clusters, except for MEM and UNC which cluster together (Fig. 2D). This supports the notion that the evaluated concentration methods reproducibly produce different outcomes for different taxa.

To quantify the enrichment effect of each method on detected viral taxa, we used the mean read count per viral strain detected in the three unconcentrated samples as a reference point. Ratios of the read counts per concentration method were compared to the unconcentrated mean to obtain an enrichment factor. PEG outperformed all methods, followed closely by VDC, in enriching read counts per virus with median enrichment factors of 9.72 and 7.63 respectively. MEM often returned lower read counts than the unconcentrated samples (median 0.37), suggesting a major reduction in the data quality of many viral taxa (Fig. 2E).

3.3. Assessing taxonomic or physical attribute biases

It would be highly beneficial to identify specific biases associated with a particular method to unveil the limitations of each method. To this end, we sought to identify taxonomic or physical attribute biases associated with method performance. We observed that the method performance was highly impacted by taxonomy. PEG was able to significantly improve enrichment in all but Papillomaviridae and was either significantly better or not significantly worse than all methods for all taxa outside of Adenoviridae where VDC performed best. MEM was the only method to perform significantly worse than the unconcentrated values across several viral families (Astroviridae, Caliciviridae, and Orthomyxoviridae). The highest enrichments were observed in members of Coronaviridae in VDC and PEG concentrates with an estimated fold enrichment of 44 and 37 respectively, supporting their utility during the SARS-CoV-2 pandemic response. Interestingly, no method significantly improved Papillomaviridae enrichment even though the morphologically similar Polyomaviridae showed moderate enrichment performance in UF, VDC, and PEG concentrates. Altogether these data indicate a substantial effect of taxonomy of method performance (Fig. 3).

While taxonomy should implicitly capture the morphological

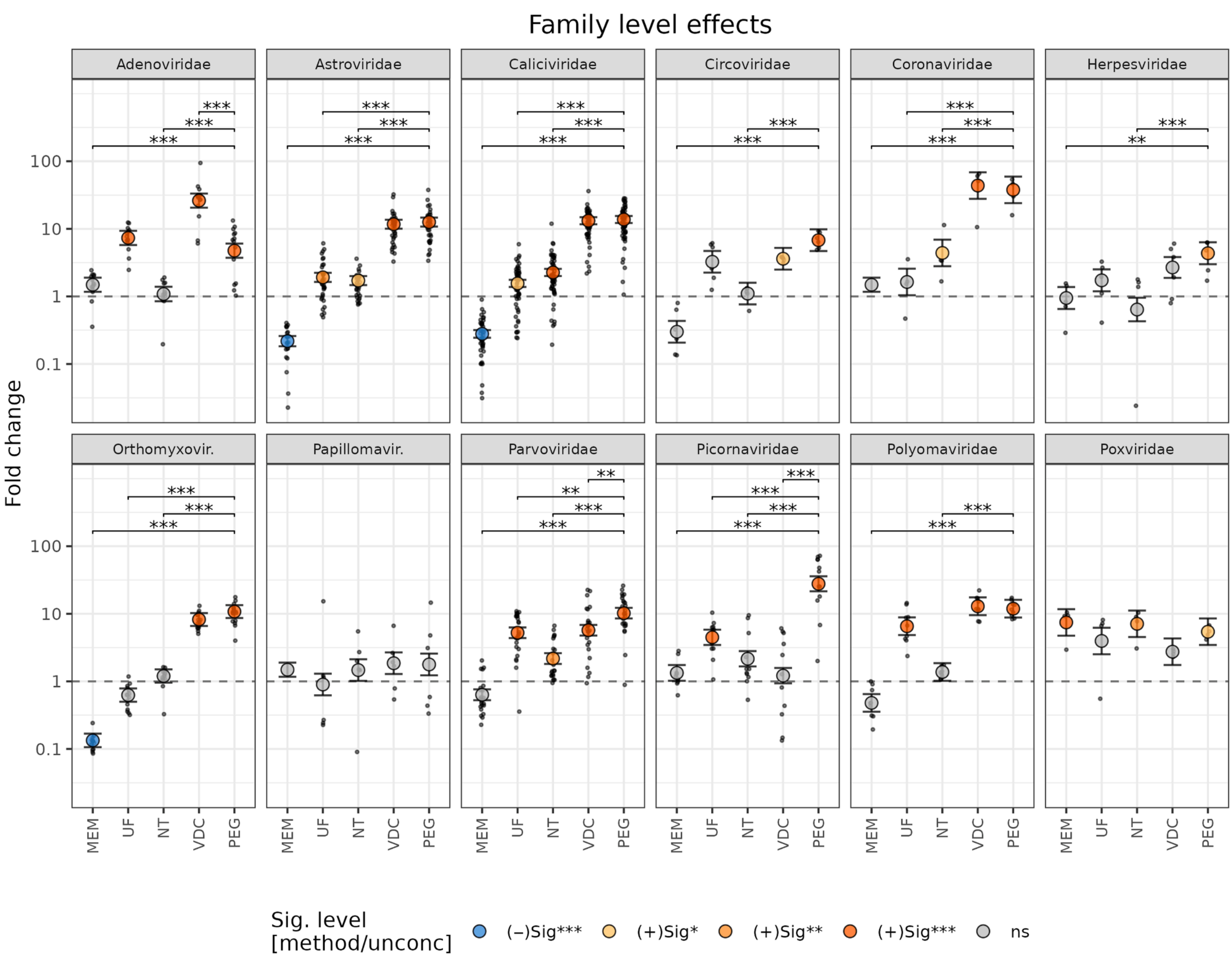

Fig. 3. Effects of concentration method on viral enrichment and diversity. Modelling results of enrichment factors with points representing model estimates, point color indicating a significant enrichment compared to the unconcentrated values, and significance brackets showing significant difference in enrichment between methods within a family. Within-family significance is shown only relative to PEG (for visualization purposes). Jitter points show the raw ratio of method to reference read count per strain detected and error bars show the 95 % confidence interval. Significance codes are as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

variations between viruses, we also directly investigated the effects of the physical attributes of each virus collected at the genus level to understand if there were sizable biases associated with a particular method. PEG, in addition to the overall highest enrichment, showed a very stable performance across all characteristics, only differing significantly in its improved enrichment of RNA viruses over DNA viruses. VDC was slightly more variable with an improved enrichment in medium sized particles and decreased performance with large particles (Suppl. Fig. S7).

3.4. Investigating captured nucleotide diversity

For evolutionary and epidemiological applications of deep wastewater sequencing it is often important to understand not only how many viral taxa are identified, but also how much diversity can be captured by each method for each taxon. An effective concentration method should capture and concentrate more viral particles per taxa, increasing the chances to capture more representative nucleotide diversity within a particular viral sequence. Calculation and pairwise comparison of the diversity metric π (as defined in Nelson and Hughes, 2015 equation 4) revealed that PEG significantly increased captured diversity over all methods except for VDC, where no significant difference was observed. To maintain comparability, this metric was only compared across the sequences detected in both methods being compared (Fig. 4).

3.5. Comparing best-in-class methods

Finally, we evaluated how reproducible the trends we observed were across different WWTPs and time points. We observed a reproducible method bias across WWTPs where viruses that tended to yield more reads (considered better enriched) for PEG or VDC did so repeatedly across varying WW matrices. The lowest correlation was at the strain level ($R = 0.53$), which can be expected given the increase in stochastic effects due to read assignments and presence at strain level but increases to 0.68 at the family level indicating that this trend holds across repeat measurements (Fig. 5A-D).

Estimates of method bias from a negative binomial model fit on the read counts per viral strain for each method were reproducible across WWTPs with little difference in the methods performance at the family level. Notably, only taxa within Picornaviridae were significantly enriched in PEG with >10-fold more reads assigned. Adenoviridae subtaxa tended to be more enriched in VDC in all WWTPs. However, after controlling for multiple testing, no significance was confidently assigned to this interaction (Fig. 5E).

4. Discussion

This study aimed to expand upon previous comparisons of wastewater concentration methods to optimize the accessible information potential for comprehensive wastewater viromics. Comparison of polyethylene glycol precipitation (PEG), vacuum-based direct capture (VDC), NanoTrap (NT), membrane adsorption (MEM), and ultrafiltration (UF) revealed substantial differences between the resulting viral sequences found in the concentrated extracts. dPCR can provide a more direct measurement of concentration efficiency, and is a fast, cheap, and sensitive method for quantifying viral nucleic acids. However, it is limited by prior sequence knowledge, sensitivity to mutation, and the number of targets that can be measured in a single run. Our dPCR data showed that VDC performed best overall, trailed by NT and PEG. Support for VDC's high performance in dPCR data is shown in an independent study showing that this method outperformed PEG, UF and MEM in SARS-CoV-2 recovery (Dimitrakopoulos et al. 2022).

However, it is often important to cast a much wider net when profiling viruses from wastewater to identify and investigate multiple taxa, including taxa for which little *a priori* knowledge is available, or when targeted methods are hampered by sequence variability within the target species. Virome-scale sequencing provides the advantage over dPCR in such cases allowing for both detection and probing of the genomic information across all detectable taxa present in a sample. Given the compositional nature of sequencing data, the detection success and accompanying depth of genomic information are heavily affected by the specificity of the method for viral particle enrichment. Method choice highly impacts the ability to detect viral taxa as exemplified in Fig. 2C and Supplemental Figure S9, which indicate the level of false negatives an analyst may encounter when using a sub-optimal method. While dPCR metrics provided a rough predictor of sequencing performance, it was necessary to consider the 16S rRNA gene measurement to account for background sequence contamination (Fig. 1B). We observed large differences in the bacterial content per method suggesting variability in the virus specificity. This greatly hindered the sensitivity for detecting viral sequences. The impact of high bacterial content on sequencing success is exemplified in NT samples where viral

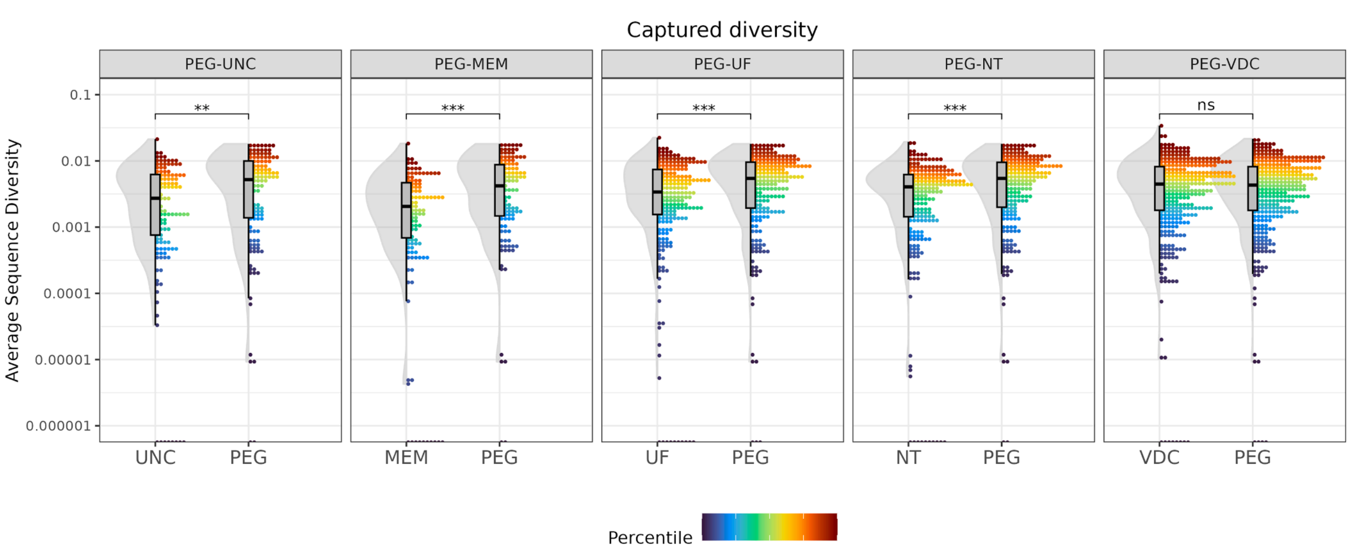

Fig. 4. Captured nucleotide diversity. Average per sequence nucleotide diversity of each method compared to PEG. Significance was determined with the Kruskal-Wallis rank-sum test. Dots indicate a single reference sequence and color is scaled by within-sample percentile. Significance codes are as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

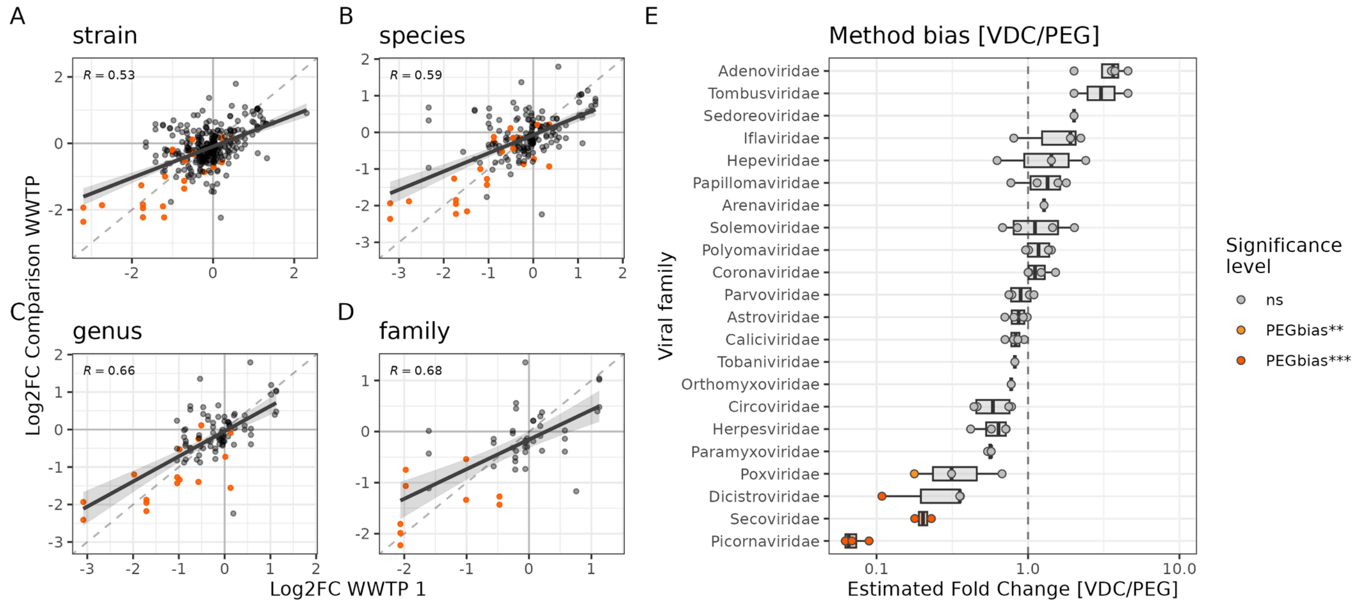

Fig. 5. Comparison of PEG and VDC across multiple WWTPs. (A-D) Reproducibility of method bias across WWTPs. The x-axis values are log_2 read ratios [VDC:PEG] in the original experiment and the y-axis values are log_2 read ratios for all three additional WWTPs tested. Point color indicates strains within each significantly biased family from the model output in panel E and the dashed line represents the perfect correlation. Correlation coefficients were computed from a linear regression through the log_2 transformed dataset. (E) Marginal effects of modelled method bias [VDC:PEG] at the family level in each WWTP. Points represent a bias estimate at a single WWTP and color represents the significance of the bias (no significant VDC biases were observed).

quantifications alone were poor indicators of the downstream sequencing performance. NT had the second highest mean performance from dPCR data but the second lowest species richness and median enrichment factors in the sequencing data likely due to the high background bacterial sequence content (Fig. 1, 2C, 2E).

Overall, we obtained several lines of evidence indicating that PEG and VDC are effective methods for concentrating a diverse set of viruses independent of physiochemical properties. PEG, VDC, and UF methods all had an initial centrifugation step to remove large particles from the sample while MEM and NT both used direct inputs without the initial centrifugation. PEG showed the highest viral specificity, which may be attributed to bacterial removal during the initial centrifugation step (Hjelmsø et al. 2017). Additionally, in support of our findings, PEG was shown to better capture low abundance viral targets over NT in sequencing data (Farkas et al., 2024). The exact mechanism of action for PEG remains poorly characterized though the prevailing hypothesis is that virus particles are aggregated in the extra-polymer space and precipitated out after exceeding their solubility (Ingham 1990). In support of our findings, Yamamoto et al. (1970) found that PEG precipitation efficiency is independent of viral concentration, pH, and ionic strength and Atha and Ingham (1981) showed that PEG has very little interaction with proteins supporting a lack of discriminatory features where performance bias can arise.

VDC is in principle a large-scale extraction and purification and relies on directly capturing nucleic acids on a silica membrane so there is not an obvious avenue for large bias introduction across taxonomy or physical particle attributes. Given this, it is surprising to observe such a high virus to bacteria ratio though likely the initial centrifugation influenced the observed viral specificity. The mechanism working via adsorption of nucleic acids, rather than adsorption of the viral particles as seen with MEM could explain the increased performance given the more variable nature of electrostatic charges between different virions.

UF relies solely on size exclusion by allowing liquids to pass through the 30 kDa filter but retaining everything else and reducing the overall volume. Poor overall performance from UF could be attributed to viral adsorption onto the membrane, which may also leave room for the preferential adsorption of certain particle types. Co-concentration of inhibitors present in the samples could also affect the assays used to produce the quantifications and sequencing results, and though inhibitor load was not explicitly investigated here, the enrichment rate for RNA viruses was significantly less than that of DNA viruses in UF and reverse transcription is a step sensitive to the presence of inhibitors (Suppl. Fig. S7).

MEM relies on the charge properties of viruses requiring a negatively charged particle to bind via $MgCl_2$ mediated salt-bridging to the electronegative membrane (Cashdollar and Wymer, 2013), which can be impacted by properties of the wastewater matrix (pH, turbidity). The limited performance of MEM relative to the other methods found in the present study is supported by dPCR-based results from Antkiewicz et al. (2024) which showed VDC, NT, and PEG outperform MEM methods when quantifying SARS-CoV-2 and bovine coronavirus.

NT relies on chemical affinity capture of viral particles and has been shown to vary in capture rate based on virus size and envelope composition (Lin et al. 2020). NT has also been shown previously to detect fewer species than PEG following hybrid-capture sequencing (Jiang et al. 2024). While our data suggests NT to reliably quantify a range of dPCR targets, it falls short in the sequencing results. The NT Microbiome A Particles used in this study enrich viral particles as well as certain bacteria expanding the range of targets for concentration, at the expense of selectivity which can be beneficial or detrimental depending on the desired use-case. A benefit to NT is the shorter hands-on time and ease of automation allowing this method to be readily scaled.

We could not identify systematic differences between the two top method's performances outside of the enrichment of Picornaviridae in PEG. This viral family contains enteroviruses which are of particular relevance, given that many WBE programs screen for poliovirus sequences (Brouwer et al., 2018; Whitehouse et al., 2025). However, the PEG bias seems to be driven by a subset of genera within Picornaviridae, namely *Kobuvirus* and *Salivirus*, while *Enterovirus* bias is largely unaffected by method choice (Suppl. Fig. S8).

This study has several limitations. We employed protocols that were previously shown to perform well in published literature. However, it is likely that modifications to these protocols could boost or hinder performance. For instance, Farkas et al. (2022) showed that addition of beef extract-sodium nitrate elution to PEG precipitation greatly improved viral recovery. Even though we followed protocols for MEM similar to those used in Ahmed et al. (2020) and LaTurner et al. (2021), our findings contradict the high levels of recovery observed in MEM and low recoveries from PEG. Furthermore, there are additional methods to concentrate viruses which were not included as it was out of scope of this study to include all potential methods and variations.

Linking the observed biases to physicochemical properties rather than the taxonomic identity of detected viruses would allow more generalizable results as the physical characteristics of the virus directly interact with the WW matrix and concentration method's mechanism of action. However, due to inherent class imbalances these properties are difficult to adequately model leaving taxonomy as a more balanced alternative. In general, we did not observe much variation on the genus level though this observation is only meaningful if a representative portion of the genus was captured in our sample set. Despite this limitation, PEG and VDC were not outperformed by any method across all observed taxonomic groups supporting the notion that either method could be the initial choice for profiling novel viruses, unless alternative approaches are systematically evaluated and demonstrated to be superior. As research and public health interest in WBE continues to grow, the need for standardized protocols and data reporting becomes paramount to ensure a coherent and comparable data ecosystem. Further work should extend these investigations into the effects of physical viral attributes on concentration efficiency to provide robust recommendations for methodology when considering future targets of interest.

5. Conclusion

In summary, we investigated the ability of five widely used WW concentration methods to effectively concentrate a broad range of viruses from different wastewater sources. dPCR was used to assess raw quantifications and performances of each method, while deep hybrid-capture sequencing was used to examine performance across the detectable human virome in WW. This study aimed to augment previous work evaluating the suitability of concentration methods for recovering viral particles from WW by expanding the range of viral targets across diverse taxonomy and physical attributes. We explore unique readouts, including captured nucleotide diversity, and detail analysis frameworks to thoroughly explore the strengths and limits of each method. We found PEG and VDC to be the best-in-class methods for viral concentration providing high relative viral concentrations and technical stability across a broad range of viral taxa, and the highest capture of nucleotide diversity within detected viral sequences. We observed reproducible trends with little systematic differences between PEG and VDC across independent WWTP sampling, highlighting both methods to be viable choices for virome-scale concentration in support of WBE.

CRediT authorship contribution statement

Matthew Thornton: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Gabriela Eder:** Formal analysis, Data curation, Writing – review & editing, Methodology, Investigation. **Fabian Amman:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anastasiia Pantielieieva:** Data curation. **Julia Vierheilig:** Formal analysis, Data curation, Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Andreas Bergthaler:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Andreas Bergthaler reports financial support was provided by Austrian Science Fund. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Professor Regina Sommer for kindly providing the PhiX174 and MS2 phage stocks, the Biomedical Sequencing Facility of the Research Center for Molecular Medicine for the sequencing of all libraries in this study, Helene Loges for conducting the 16S rRNA gene quantifications, and the WWTP operators for providing samples. The work was financially supported by the Austrian Science Fund, Cluster of Excellence COE7 and the FFG KIRAS project 909310 granted to AB.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2025.123985.

Data availability

All code used in the analysis and figure production can be viewed on Github (https://github.com/thorntonmr/comparative_ww_virome_analysis_manuscript). The data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB86516 (https://www.ebi.ac.uk/ena/browser/view/PRJEB86516).

References

Adams, C., Bias, M., Welsh, RM., Webb, J., Reese, H., Delgado, S., Person, J., West, R., Shin, S., Kirby, A., 2024. The National Wastewater Surveillance System (NWSS): from inception to widespread coverage, 2020–2022, United States. Sci. Total Environ. 924 (May), 171566. https://doi.org/10.1016/j.scitotenv.2024.171566.

Ahmed, W., Bertsch, PM., Bivins, A., Bibby, K., Farkas, K., Gathercole, A., Haramoto, E., et al., 2020. Comparison of virus concentration methods for the RT-qPCR-based recovery of Murine Hepatitis virus, a surrogate for SARS-CoV-2 from untreated wastewater. Sci. Total Environ. 739 (October), 139960. https://doi.org/10.1016/j.scitotenv.2020.139960.

Ahmed, W., Bivins, A., Stephens, M., Metcalfe, S., Smith, WJ.M., Sirikanchana, K., Kitajima, M., Simpson, SL., 2023. Occurrence of multiple Respiratory viruses in wastewater in Queensland, Australia: potential for community disease surveillance. Sci. Total Environ. 864 (March), 161023. https://doi.org/10.1016/j.scitotenv.2022.161023.

Amman, F., Markt, R., Endler, L., Hupfauf, S., Agerer, B., Schedl, A., Richter, L., et al., 2022. Viral variant-resolved wastewater surveillance of SARS-CoV-2 at national scale. Nat. Biotechnol. 40 (12), 1814–1822. https://doi.org/10.1038/s41587-022-01387-y.

Antkiewicz, S., Dagmara, KH.J, Roguet, A., Pilch, HE., Fahney, RB., Mullen, PA., Knuth, GN., et al., 2024. Wastewater-based protocols for SARS-CoV-2: insights into virus concentration, extraction, and quantitation methods from two years of public health surveillance. Environ. Sci. 10 (8), 1766–1784. https://doi.org/10.1039/D3EW00958K.

Atha, D.H., Ingham, K.C., 1981. Mechanism of precipitation of proteins by polyethylene glycols. Analysis in terms of excluded volume. J. Biol. Chem. 256 (23), 12108–12117. https://doi.org/10.1016/S0021-9258(18)43240-1.

Barril, P.A., Pianciola, L.A., Mazzeo, M., Ousset, M.J., Jaureguiberry, M.V., Alessandrello, M., Sánchez, G., Oteiza, J.M., 2021. Evaluation of viral concentration methods for SARS-CoV-2 recovery from wastewaters. Sci. Total Environ. 756 (February), 144105. https://doi.org/10.1016/j.scitotenv.2020.144105.

Brouwer, AF., Eisenberg, JN.S., Pomeroy, CD., Shulman, LM., Hindiyeh, M., Manor, Y., Grotto, I., Koopman, JS., Eisenberg, MC., 2018. Epidemiology of the silent polio outbreak in Rahat, Israel, based on modeling of environmental surveillance data. Pro. Natl. Acad. Sci. 115 (45), E10625–E10633. https://doi.org/10.1073/pnas.1808798115.

Cashdollar, J.L., Wymer, L., 2013. Methods for primary concentration of viruses from water samples: a review and meta-analysis of recent studies. J. Appl. Microbiol. 115 (1), 1–11. https://doi.org/10.1111/jam.12143.

Cheshomi, N., Alum, A., Smith, MF., Lim, ES., Conroy-Ben, O., Abbaszadegan, M., 2024. Viral concentration method biases in the detection of Viral profiles in wastewater. Appl. Environ. Microbiol. 91 (1), e01339. https://doi.org/10.1128/aem.01339-24. -24.

Choi, PM., Tscharke, BJ., Donner, E., O'Brien, JW., Grant, SC., Kaserzon, SL., Mackie, R., et al., 2018. Wastewater-based epidemiology biomarkers: past, present and future. TrAC Trends Anal. Chem. 105 (August), 453–469. https://doi.org/10.1016/j.trac.2018.06.004.

Dimitrakopoulos, L., Kontou, A., Strati, A., Galani, A., Kostakis, M., Kapes, V., Lianidou, E., Thomaidis, N., Markou, A., 2022. Evaluation of viral concentration and extraction methods for SARS-CoV-2 recovery from wastewater using droplet digital and quantitative RT-PCR. Case Stud. Chem. Environ. Eng. 6 (December), 100224. https://doi.org/10.1016/j.cscee.2022.100224.

Farkas, K., Hillary, LS., Thorpe, J., Walker, DI., Lowther, JA., McDonald, JE., Malham, SK., Jones, DL., 2021. Concentration and quantification of SARS-CoV-2 RNA in wastewater using polyethylene glycol-based Concentration and qRT-PCR. Methods Protoc. 4 (1), 17. https://doi.org/10.3390/mps4010017.

Farkas, K., Pellett, C., Alex-Sanders, N., Bridgman, MT.P., Corbishley, A., Grimsley, JM. S., Kasprzyk-Hordern, B., et al., 2022. Comparative assessment of filtration- and precipitation-based methods for the concentration of SARS-CoV-2 and other viruses from wastewater. Microbiol. Spectr. 10 (4), e01102–e01122. https://doi.org/10.1128/spectrum.01102-22.

Farkas, K., Kevill, J.L., Williams, R.C., Pântea, I., Ridding, N., Lambert-Slosarska, K., Woodhall, N., et al., 2024. Comparative assessment of nanotrap and polyethylene glycol-based virus concentration in wastewater samples. FEMS. Microbes. 5 (January), xtae007. https://doi.org/10.1093/femsmc/xtae007.

Girón-Guzmán, I., Díaz-Reolid, A., Cuevas-Ferrando, E., Falcó, I., Cano-Jiménez, P., Comas, I., Pérez-Cataluña, A., Sánchez, G., 2023. Evaluation of two different concentration methods for surveillance of Human viruses in sewage and their effects on SARS-CoV-2 sequencing. Sci. Total Environ. 862 (March), 160914. https://doi.org/10.1016/j.scitotenv.2022.160914.

Gracia-Lor, E., Castiglioni, S., Bade, R., Been, F., Castrignanò, E., Covaci, A., González-Mariño, I., et al., 2017. Measuring biomarkers in wastewater as a new source of epidemiological information: current State and future perspectives. Environ. Int. 99 (February), 131–150. https://doi.org/10.1016/j.envint.2016.12.016.

Hjelmsø, M.H., Hellmér, M., Fernandez-Cassi, X., Timoneda, N., Lukjancenko, O., Seidel, M., Elsässer, D., et al., 2017. Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing. PLoS. One 12 (1), e0170199. https://doi.org/10.1371/journal.pone.0170199.

Huizer, M., Laak, TL., Voogt, P., Wezel, AP., 2021. Wastewater-based epidemiology for illicit drugs: a critical review on global data. Water Res. 207 (December), 117789. https://doi.org/10.1016/j.watres.2021.117789.

Ingham, Kenneth C., 1990. Precipitation of proteins with polyethylene glycol. In: Deutscher, Murray P. (Ed.), Volume 182: Guide to Protein Purification, Methods in Enzymology, 182. Academic Press, pp. 301–306. https://doi.org/10.1016/0076-6879(90)82025-W edited byGuide to Protein Purification.

Jafferali, M.H., Khatami, K., Atasoy, M., Birgersson, M., Williams, C., Cetecioglu, Z., 2021. Benchmarking virus concentration methods for quantification of SARS-CoV-2 in raw wastewater. Sci. Total Environ. 755 (February), 142939. https://doi.org/10.1016/j.scitotenv.2020.142939.

Jahn, K., Dreifuss, D., Topolsky, I., Kull, A., Ganesanandamoorthy, P., Fernandez-Cassi, X., Bänziger, C., et al., 2022. Early detection and surveillance of SARS-CoV-2 genomic variants in wastewater using COJAC. Nat. Microbiol. 7 (8), 1151–1160. https://doi.org/10.1038/s41564-022-01185-x.

Jiang, M., Wang, AL.W., Be, NA., Mulakken, N., Nelson, KL., Kantor, RS., 2024. Evaluation of the impact of concentration and extraction methods on the targeted sequencing of Human viruses from wastewater. Environ. Sci. Technol. 58 (19), 8239–8250. https://doi.org/10.1021/acs.est.4c00580.

Karthikeyan, S., Levy, JI., Hoff, PDe, Humphrey, G., Birmingham, A., Jepsen, K., Farmer, S., et al., 2022. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. Nature 609 (7925), 101–108. https://doi.org/10.1038/s41586-022-05049-6.

Kilaru, P., Hill, D., Anderson, K., Collins, M.B., Green, H., Kmush, B.L., Larsen, D.A., 2023. Wastewater surveillance for infectious disease: a systematic review. Am. J. Epidemiol. 192 (2), 305–322. https://doi.org/10.1093/aje/kwac175.

Krogsgaard, L.W., Benedetti, G., Gudde, A., Richter, S.R., Rasmussen, L.D., Midgley, S.E., Qvesel, A.G., Nauta, M., Bahrenscheer, N.S., von Kappelgaard, McManus, O., 2024. Results from the SARS-CoV-2 wastewater-based surveillance system in Denmark, July 2021 to June 2022. Water Res. 252, 121223. https://doi.org/10.1016/j.watres2024.121223.

LaTurner, ZW., Zong, DM., Kalvapalle, P., Gamas, K.R., Terwilliger, A., Crosby, T., Ali, P., et al., 2021. Evaluating recovery, cost, and throughput of different concentration methods for SARS-CoV-2 wastewater-based epidemiology. Water Res. 197 (June), 117043. https://doi.org/10.1016/j.watres.2021.117043.

Li, Y., Miyani, B., Childs, KL., Shiu, S.-H., Xagoraraki, I., 2024. Effect of wastewater collection and concentration methods on assessment of viral diversity. Sci. Total Environ. 908, 168128. https://doi.org/10.1016/j.scitotenv.2023.168128. January 15.

Lin, S.-C., Carey, BD., Callahan, V., Lee, Ji-H, Bracci, N., Patnaik, A., Smith, AK., Narayanan, A., Lepene, B., Kehn-Hall, K., 2020. Use of nanotrap particles for the capture and enrichment of Zika, Chikungunya and dengue viruses in urine. PLoS. One 15 (1), e0227058. https://doi.org/10.1371/journal.pone.0227058.

Melvin, RG., Hendrickson, EN., Chaudhry, N., Georgewill, O., Freese, R., Schacker, TW., Simmons, GE., 2021. A novel wastewater-based epidemiology indexing method predicts SARS-CoV-2 disease prevalence across treatment facilities in metropolitan and regional populations. Sci. Rep. 11 (1), 21368. https://doi.org/10.1038/s41598-021-00853-y.

Nelson, CW., Hughes, AL., 2015. Within-host nucleotide diversity of virus populations: insights from next-generation sequencing. Infect. Genet. Evol. 30 (March), 1–7. https://doi.org/10.1016/j.meegid.2014.11.026.

Oghuan, J., Chavarria, C., Vanderwal, SR., Gitter, A., Ojaruega, A.A., Monserrat, C., Bauer, CX., et al., 2023. Wastewater analysis of Mpox virus in a City with low prevalence of mpox disease: an environmental surveillance study. Lancet Regional Health 28 (December). https://doi.org/10.1016/j.lana.2023.100639.

Pino, NJ., Rodriguez, DC., Cano, L.C., Alejandra, R., 2021. Detection of SARS-CoV-2 in wastewater is influenced by sampling time, concentration method, and target analyzed. J. Water Health 19 (5), 775–784. https://doi.org/10.2166/wh.2021.133.

Polo, D., Quintela-Baluja, M., Corbishley, A., Jones, DL., Singer, AC., Graham, DW., Romalde, JL., 2020. Making waves: wastewater-based epidemiology for COVID-19 – Approaches and challenges for surveillance and prediction. Water Res. 186 (November), 116404. https://doi.org/10.1016/j.watres.2020.116404.

Radu, E., Masseron, A., Amman, F., Schedl, A., Agerer, B., Endler, L., Penz, T., et al., 2022. Emergence of SARS-CoV-2 alpha lineage and its correlation with quantitative wastewater-based epidemiology data. Water Res. 215 (May), 118257. https://doi.org/10.1016/j.watres.2022.118257.

Salvo, M., Moller, A., Alvareda, E., Gamazo, P., Colina, R., Victoria, M., 2021. Evaluation of low-cost viral concentration methods in wastewaters: implications for SARS-CoV-2 pandemic surveillances. J. Virol. Methods 297 (November), 114249. https://doi.org/10.1016/j.jviromet.2021.114249.

Sherchan, SP., Solomon, T., Oladele, I., Daniel, N., Ocean, T., 2023. Wastewater surveillance of Mpox virus in Baltimore. Sci. Total Environ. 891 (September), 164414. https://doi.org/10.1016/j.scitotenv.2023.164414.

Tambini, G., Andrus, JK., Marques, E., Boshell, J., Pallansch, M., Quadros, OA., Kew, O., 1993. Direct detection of wild poliovirus circulation by stool surveys of healthy children and analysis of community wastewater. J. Infect. Dis. 168 (6), 1510–1514. https://doi.org/10.1093/infdis/168.6.1510.

Tisza, M., Javornik Cregeen, S., Avadhanula, V., Zhang, P., Ayvaz, T., Feliz, K., Hoffman, KL., et al., 2023. Wastewater sequencing reveals community and variant dynamics of the collective Human virome. Nat. Commun. 14 (1), 6878. https://doi.org/10.1038/s41467-023-42064-1.

Vitale, D., Morales Suárez-Varela, M., Picó, Y., 2021. Wastewater-based epidemiology, a tool to bridge biomarkers of exposure, contaminants, and Human health. Curr. Opin. Environ. Sci. Health 20 (April), 100229. https://doi.org/10.1016/j.coesh.2021.100229.

Whitehouse, E.R., Gerloff, N., English, R., Reckling, S.K., Alazawi, M.A., Fuschino, M., St George, K., et al. n.d. "Wastewater Surveillance for poliovirus in selected jurisdictions, United States, 2022–2023 - volume 30, number 11—November 2024 - Emerging Infectious Diseases Journal - CDC." Accessed February 3, 2025. https://doi.org/10.3201/eid3011.240771.

Wolfe, MK., Duong, D., Bakker, KM., Ammerman, M., Mortenson, L., Hughes, B., Arts, P., et al., 2022. Wastewater-based detection of two influenza outbreaks. Environ. Sci. Technol. Lett. 9 (8), 687–692. https://doi.org/10.1021/acs.estlett.2c00350.

Yamamoto, KR., Alberts, BM., Benzinger, R., Lawhorne, L., Treiber, G., 1970. Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. Virology 40 (3), 734–744. https://doi.org/10.1016/0042-6822(70)90218-7.

Zheng, X., Deng, Yu, Xu, X., Li, S., Zhang, Y., Ding, J., On, H.Y., et al., 2022. Comparison of virus concentration methods and RNA extraction methods for SARS-CoV-2 wastewater surveillance. Sci. Total Environ. 824 (June), 153687. https://doi.org/10.1016/j.scitotenv.2022.153687.