ACS ES&T Water

pubs.acs.org/estwater

Article

Robust Performance of SARS-CoV-2 Whole-Genome Sequencing from Wastewater with a Nonselective Virus Concentration Method

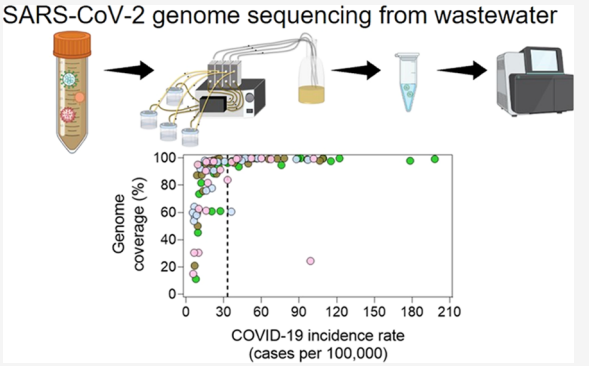
Emily Segelhurst, Jonathan E. Bard, Annemarie N. Pillsbury, Md Mahbubul Alam, Natalie A. Lamb, Chonglin Zhu, Alyssa Pohlman, Amanda Boccolucci, Jamaal Emerson, Brandon J. Marzullo, Donald A. Yergeau, Norma J. Nowak, Ian M. Bradley, Jennifer A. Surtees, and Yinyin Ye*

Cite This: *ACS EST Water* 2023, 3, 954–962

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The sequencing of human virus genomes from wastewater samples is an efficient method for tracking viral transmission and evolution at the community level. However, this requires the recovery of viral nucleic acids of high quality. We developed a reusable tangential-flow filtration system to concentrate and purify viruses from wastewater for genome sequencing. A pilot study was conducted with 94 wastewater samples from four local sewersheds, from which viral nucleic acids were extracted, and the whole genome of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was sequenced using the ARTIC V4.0 primers. Our method yielded a high probability (0.9) of recovering complete or near-complete SARS-CoV-2 genomes (>90% coverage at 10× depth) from wastewater when the COVID-19 incidence rate exceeded 33 cases per 100 000 people. The relative abundances of sequenced SARS-CoV-2 variants followed the trends observed from patient-derived samples. We also identified SARS-CoV-2 lineages in wastewater that were underrepresented or not present in the clinical whole-genome sequencing data. The developed tangential-flow filtration system can be easily adopted for the sequencing of other viruses in wastewater, particularly those at low concentrations.



KEYWORDS: *SARS-CoV-2, COVID-19, wastewater genome sequencing, tangential-flow filtration, wastewater-based epidemiology*

INTRODUCTION

The global spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has facilitated the emergence of genome mutations, resulting in new lineages that further threaten public health.[1] Whole-genome sequencing (WGS) of clinical samples is a powerful method for tracking the spread of various SARS-CoV-2 lineages,[2] but this can be expensive, slow, and subject to biased sampling. Furthermore, the frequency of clinical testing has declined, aided by the availability of rapid at-home testing kits. Sequencing SARS-CoV-2 genomes in wastewater samples circumvents the issues as a high-throughput and more comprehensive strategy that monitors a larger and unbiased portion of the population.[3−6] Successful sequencing of SARS-CoV-2 genomes in wastewater samples relies on the depth (the number of times a nucleotide is sequenced)[7] and breadth of genome coverage (hereafter, coverage; the percentage of nucleotide positions sequenced to a given depth).[7] Because SARS-CoV-2 lineages differ by only a few mutations, maximum depth and coverage are needed to quantify lineage abundance and detect low-frequency mutations.[8,9]

Various virus concentration methods have been applied to perform polymerase chain reaction (PCR)-based detection of viral genes from wastewater.[10−13] However, the nucleic acids recovered with these methods may not be suitable for genome sequencing—PCR assays typically target <1% of the genome. Inhibitors of the enzymes used for sequencing, such as humic acids and nucleic acids from prokaryotes and eukaryotes, are likely to reduce the genome coverage of SARS-CoV-2. Indeed, near-complete (>90%) SARS-CoV-2 genomes are not recovered from most samples (>85%) when aluminum hydroxide is used to directly precipitate viruses without separating viruses from other microorganisms in wastewater.[14] The removal of wastewater solids before the Amicon/Centricon ultracentrifugal concentration[15,16] and electronegative membrane filtration[17] can improve sequencing success, but the quality of the extracted genomes can vary, and random sequencing failures can occur.[18] These previous results

Received: September 15, 2022
Revised: January 1, 2023
Accepted: February 21, 2023
Published: March 2, 2023

suggested that concentrating virus particles with separation from chemical and biological substances might improve the genome sequence coverage from wastewater. Despite being less applied to collect longitudinal viral genome sequencing data from wastewater, a tangential-flow filtration system has been previously used to detect viral genes in wastewater.[19−21] This approach concentrates viruses and removes nonviral genomes and cell-free nucleic acids through size-exclusion membranes.[22] Moreover, due to minimum membrane fouling issues, the ultrafilter membranes can be washed for reuse to save operation costs.[23] We hypothesized that tangential-flow filtration would provide a robust, cost-effective method to recover viral nucleic acids that are suitable for sequencing near-complete viral genomes from wastewater.

To test our hypothesis, we developed an in-house built tangential-flow filtration system followed by multiplexed amplicon sequencing of SARS-CoV-2 genome. Using this workflow, we tracked SARS-CoV-2 lineages in the wastewater samples collected from four local sewersheds over 6 months. We subsequently compared the findings with those from patient-derived clinical samples at the County level. Finally, the timing of lineage detection was compared among clinical WGS, wastewater genome sequencing, and quantitative reverse transcription polymerase chain reactions (RT-qPCR).

MATERIALS AND METHODS

Wastewater Sample Collection. Time- or flow-weighted 24-hr composite influent samples were collected every 1 or 2 weeks between October 1 and December 16, 2021, and between January 18 and April 12, 2022, from four wastewater treatment plants covering four sewersheds that serve ∼80% of the total population of Erie County (New York): Tonawanda, Kenmore−Tonawanda, Amherst, and Bird Island (Table S1 and Figure S1). Influent was collected by an autosampler every 30 min, and the samples were kept at 4 °C in high-density polyethylene bottles that were cleaned with 10% bleach. The samples were transported on ice packs and stored at 4 °C until they were processed within 12 h.

Virus Concentration from Wastewater. Wastewater samples (125 mL) were centrifuged in sterile bottles at 10 000*g* for 15 min at 4 °C to remove large wastewater solids, prokaryotic and eukaryotic cells, and other debris. The supernatant was concentrated in a Vivaflow laboratory cross-flow cassette system (Sartorius) equipped with a 30 kDa Hydrosart ultrafilter membrane (Sartorius) at a feedline flow rate of 8.5 mL/min (Figure S2). The residual liquid in the ultrafilter membrane and connected tubing was blown with air into the sample reservoir. The concentrate (∼25 mL) in the reservoir was then collected, overlayed with 5 mL of a 20% (v/v) sucrose solution, and ultracentrifuged at 100 000*g* for 45 min at 4 °C in an SW 32 Ti rotor using an Optima XE series centrifuge (Beckman Coulter). Immediately after the ultracentrifugation, the pellet was resuspended in 200 μL phosphate-buffered saline (pH 7.4; Gibco) and stored at −80 °C ∼2 weeks until nucleic acid extraction.

After use, the ultrafilter membranes were immediately flushed with ∼100 mL of 0.5 M NaOH solution preheated to 55 °C at a feedline flow rate of 8.5 mL/min. The NaOH solution was then recirculated for 20−30 min at the same flow rate. The cleaned membrane was stored in 0.5 M NaOH at 4 °C. Cleaning efficiency was verified on 6 random days by processing 125 mL of autoclaved Milli-Q water with the same concentration procedure. The absence of SARS-CoV-2 genes through washed filters was confirmed by RT-qPCR (Figure S3). Although we did not optimize the potential life span of the ultrafilter membranes, we determined that they could readily be washed and reused by this protocol at least 20 times.

Nucleic Acid Extraction and RT-qPCR. Viral nucleic acids were extracted from the concentrated samples with QIAamp viral RNA mini kits (Qiagen) and eluted in 60 μL AVE buffer (Qiagen) per the manufacturer's instructions. One-step RT-qPCR was performed to quantify the SARS-CoV-2 N gene according to the CDC N2 assay[24] and to determine the presence/absence of specific S gene mutations, including WT493-498, Q493R and Q498R, delH69/V70, and delL24/P25/P26 and A27S as variant determinants for Delta, generic Omicron, Omicron BA.1, and Omicron BA.2, respectively, following methods published previously (Table S2).[25,26] The limit of detection for the SARS-CoV-2 N gene was determined to be 1.8 gene copies/μL, and the limit of quantification was determined to be 5 gene copies/μL following a method as described previously (Figure S4).[27] A checklist of the MIQE (minimum information for publication of quantitative real-time PCR experiments) guidelines[28] is provided in Table S3. All RT-qPCR assays were conducted in duplicate on a CFX96 touch real-time PCR detection system (Bio-Rad), and the threshold cycle (C_T) value was determined using CFX Maestro Software (Version 4.0, Bio-Rad). To remove PCR inhibition, the extracted nucleic acids were diluted 5- or 10-fold in nuclease-free water before RT-qPCR analyses (Figure S5). Detailed procedures of RT-qPCR inhibition test, reaction compositions, thermocycling conditions, and determination of the limits of detection and quantification are described in the Supporting Information.

Amplicon Sequencing of SARS-CoV-2 Genomes. The quality of the extracted nucleic acids was assessed with the Agilent fragment analyzer. Amplicon sequencing of SARS-CoV-2 genomes was performed following a modified ARTIC protocol[29] using the V4.0 nCOV-2019 amplicon panel (IDT). Briefly, 8 μL of nucleic acid extracts was reverse-transcribed to cDNA with random hexamers using the Invitrogen SuperScript IV first-strand synthesis system (Thermo Fisher Scientific) according to the manufacturer's protocol. The cDNA without further dilution was amplified with two primer pools using Q5 hot-start high-fidelity 2× master mix (New England BioLabs) with the following parameters: 98 °C for 30 s, 25 cycles of 98 °C for 15 s, and 65 °C for 5 min. The resulting amplicons from both primer pools were combined. Excess primers and reagents were removed with 1× AMPure XP beads (Beckman Coulter), and the amplicons were eluted in EB buffer (Qiagen). The total eluted volume of amplicons was input to generate libraries using the NEBNext Ultra II DNA library prep kit (New England BioLabs) without fragmentation per the manufacturer's protocol. Individual samples were then barcoded and pooled for quantification using the sparQ Universal Library Quant kit (QuantaBio) for Illumina sequencing. The libraries were diluted to 10 pM and sequenced on an Illumina MiSeq instrument (V3 chemistry, PE300), with 1% PhiX as an internal control.

Bioinformatic Analysis. The Illumina sequencing output files were demultiplexed using bcl2fastq (v2.20.0.422) to convert them into FASTQ files. Initial quality control was performed through FastQ Screen.[30] Samples were then processed using the UB Genomics and Bioinformatics Core SARS-CoV-2 analysis pipeline (https://github.com/UBGBC/fastq-to-consensus). Briefly, adapters were trimmed before

https://doi.org/10.1021/acsestwater.2c00456
ACS EST Water 2023, 3, 954−962

sequencing reads were aligned to the SARS-CoV-2 reference genome (MN908947.2) using the BWA-MEM algorithm.[31] Variants, insertions, and deletions were then called using BCFtools (v.1.10.2),[32] requiring a minimum depth per nucleotide position of 10× or 50× and generating VCF file outputs along with a final consensus FASTA file for each input sample. The resulting VCF and depth mpileup files were used as input into Freyja to perform lineage composition analysis (https://github.com/andersen-lab/Freyja).[5] The coverage (--covcut) was specified at 10× or 50×, and only confirmed lineages were reported. The Freyja pipeline was selected because of its high accuracy and efficiency.[5]

Patient-Derived SARS-CoV-2 Data. The daily cases of COVID-19 during the sampling periods in the studied sewersheds were extracted from the Erie County Wastewater Monitoring Dashboard.[33] The COVID-19 incidence rates (per 100 000 population in the sewershed) were then calculated as 7-day rolling averages. The patient-derived WGS data at the County level were downloaded from the Global Initiative on Sharing All Influenza Data (GISAID) database[34] with the location as "North America/USA/New York/Erie County," collection date from "2021-10-01" to "2022-04-30", and low coverage excluded (GISAID Identifier: EPI_SET_220906wr). Most of these samples were sequenced at the University at Buffalo.[35] On each collection date, the relative abundances of SARS-CoV-2 lineages in the patient samples were calculated, and the number of patient samples that were collected for sequencing were smoothed as 7-day rolling averages.

Statistical Analysis. The probability of recovering >90% coverage of the SARS-CoV-2 genome at 10× depth at a given RT-qPCR C_T value or COVID-19 incidence rate was predicted by binary logistic regression analysis with Prism 9.2.0 (GraphPad Software). The likelihood ratio test is used to assess if the C_T value contributes significantly to predicting sequencing success, and the 95% confidence interval of the predicted probability of sequencing success was calculated based on the standard errors of the fitting coefficients.[36] To compare the mean C_T value at >90% genome coverage and the mean probability of sequencing success at high C_T values (C_T > 32) in this study with previous studies, one-way analysis of variance (ANOVA) was applied and followed by Dunnett's tests to compare the mean value in each previous study with the mean value in this study. Data were visualized with Prism 9.2.0 and the ggplot2[37] package in R version 4.2.1.[38]

Data Availability. Raw sequencing data are available in NCBI Sequence Read Archive (SRA) under the BioProject ID: PRJNA877272. Codes for analyzing SARS-CoV-2 lineage in wastewater are available at https://github.com/UBGBC/fastq-to-consensus.

RESULTS AND DISCUSSION

Coverage and Depth of SARS-CoV-2 Genomes from Wastewater. Complete or near-complete SARS-CoV-2 genomes (>90% coverage) at 10× depth were recovered from 68% (64/94) of the wastewater samples (Table S4). A binary logistic regression analysis predicted a 0.9 probability of sequencing success (>90% coverage at 10× depth) when the C_T value of the SARS-CoV-2 N gene was 31.8, corresponding to ~3500 gene copies loaded for sequencing (N = 94, likelihood ratio test p < 0.0001, Figure 1A). For comparison, at the >0.9 probability of sequencing success, the mean C_T value of the SARS-CoV-2 gene in this study was significantly higher than the mean C_T values reported in previous pilot studies that

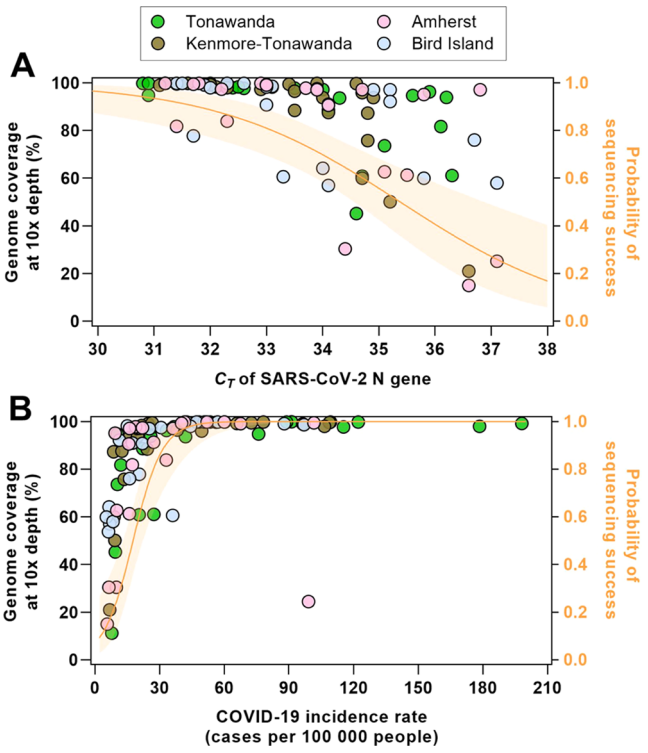


Figure 1. Coverage of the SARS-CoV-2 genome at 10× depth and the probability of sequencing success (>90% coverage at 10× depth) with regard to C_T value of the SARS-CoV-2 N gene measured with RT-qPCR (A) and the COVID-19 incidence rate (rolling 7-day average of cases per 100 000 people) (B). The probability of sequencing success was calculated according to a binary logistic regression analysis.

used mixed cellulose esters membrane[17] (Dunnett's test p < 0.0001) and aluminum hydroxide precipitation[14] (Dunnett's test p < 0.0001) to concentrate viruses and sequenced with ARTIC primers, and that applied Amicon ultrafiltration and sequenced with Nanopore native bar code kit[18] (Dunnett's test p = 0.016). Our mean C_T value was also greater than that reported in Fontenele et al.[15] but was not statistically significant (Dunnett's test p = 0.103). To assess the probability of sequencing success (>90% coverage) at higher C_T values, we conducted a binary logistic analysis with the genome coverage and its corresponding C_T reported in previous studies (Figure S6).[14,15,17,18] Our workflow demonstrated a significantly higher mean probability of sequencing success than the mean probability estimated in previous studies at C_T values greater than 32 (ANOVA p < 0.0001, all Dunnett's test p < 0.0001). Note that the variations in wastewater samples, sewersheds, target region for RT-qPCR quantification, multiplexed primers, and sequencing platforms may all contribute to the varied genome coverage. Future work will include a parallel comparison of the impact of various nonselective virus concentration methods on the reliability of sequencing viral genomes from wastewater.

Most of our wastewater samples were sequenced to an average depth of >50× (Figure S7). We noted that three regions were often sequenced at lower depths (<10×): nucleotides 21 750−22 250 in the S gene (amplicons 72 and 73) and 26 750−27 000 in the M gene (amplicon 89) (Figure S7). Similar amplicon dropout issues were previously reported for a few amplicons using the ARTIC V3 panel.[39,40] We used ARTIC V4.0 primers, which address the V3 coverage issues

https://doi.org/10.1021/acsestwater.2c00456
ACS EST Water 2023, 3, 954−962

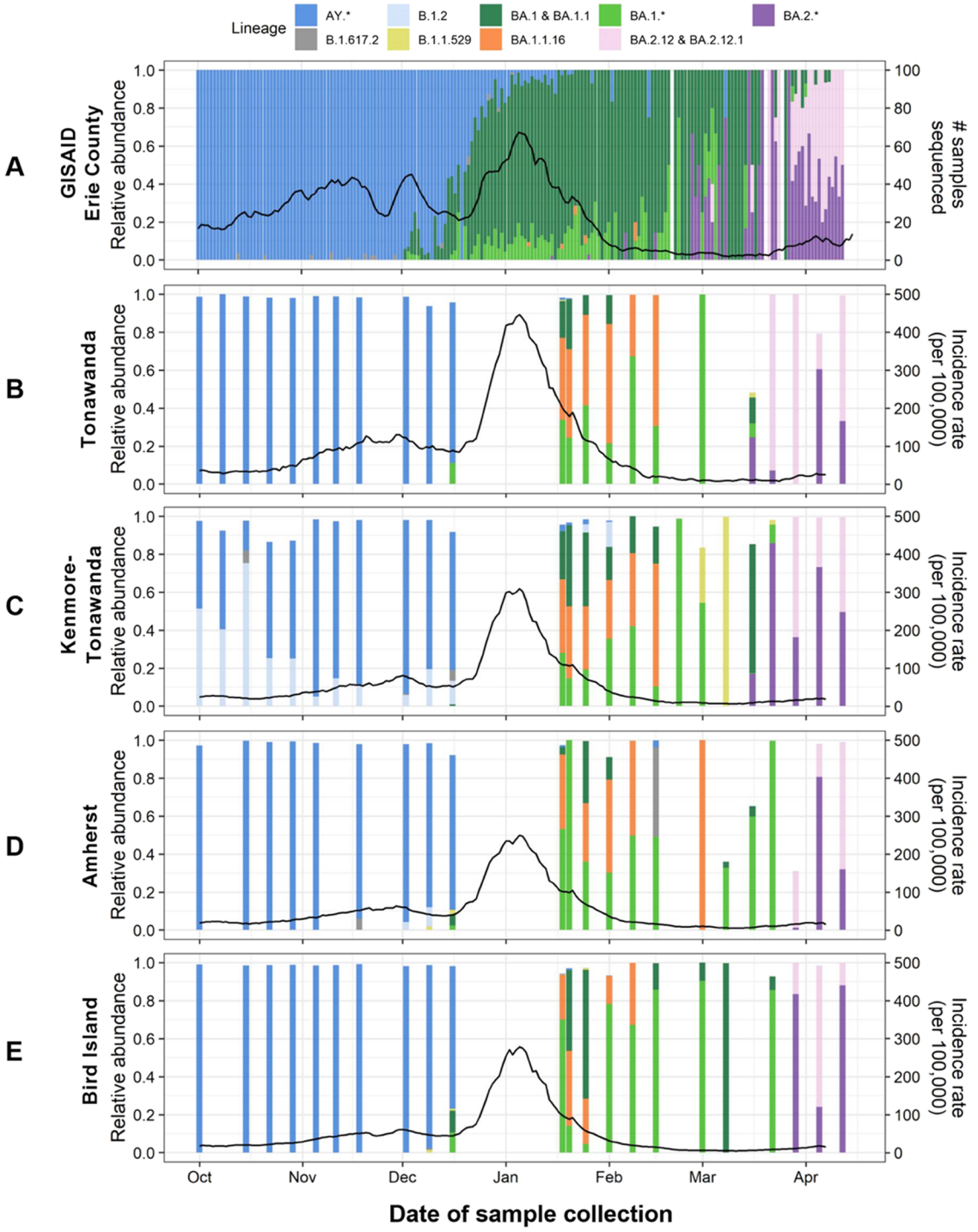


Figure 2. Distributions of dominant SARS-CoV-2 lineages estimated from Erie County clinical data deposited on GISAID (A) and wastewater sequencing data (B–E) during the sampling periods. The solid black lines represent the rolling 7-day average numbers of clinical samples collected for sequencing (A) and COVID-19 incidence rates (cases per 100 000 people) in each sewershed (B–E).

but do not include all mutations in SARS-CoV-2 Delta and Omicron lineages in the primer design.[41] Other wastewater factors may also explain the low-depth coverage within certain regions of SARS-CoV-2 genome. First, the low concentrations of the SARS-CoV-2 genome in wastewater ($<10^5$ gene copies/mL)[42] may have contributed to the lower sequencing depth. The presence of primer dimers competes for primer–template interactions, resulting in reduced amplification efficiency.[43] Also, mutations of SARS-CoV-2 genomes identified in wastewater but not in patient samples could influence primer binding at some sites.[3,6,44] Our results suggest that up-to-date primer design and optimization is critical to sequence emerging variants from wastewater.

Robustness of SARS-CoV-2 Genome Sequencing from Wastewater. *Influence of Sewersheds.* The rates of sequencing success were similar for samples collected from the four studied sewersheds: 70.8% (17/24) for Tonawanda, 72.0% (18/25) for Kenmore–Tonawanda, 72.7% (16/22) for

https://doi.org/10.1021/acsestwater.2c00456
ACS EST Water 2023, 3, 954–962

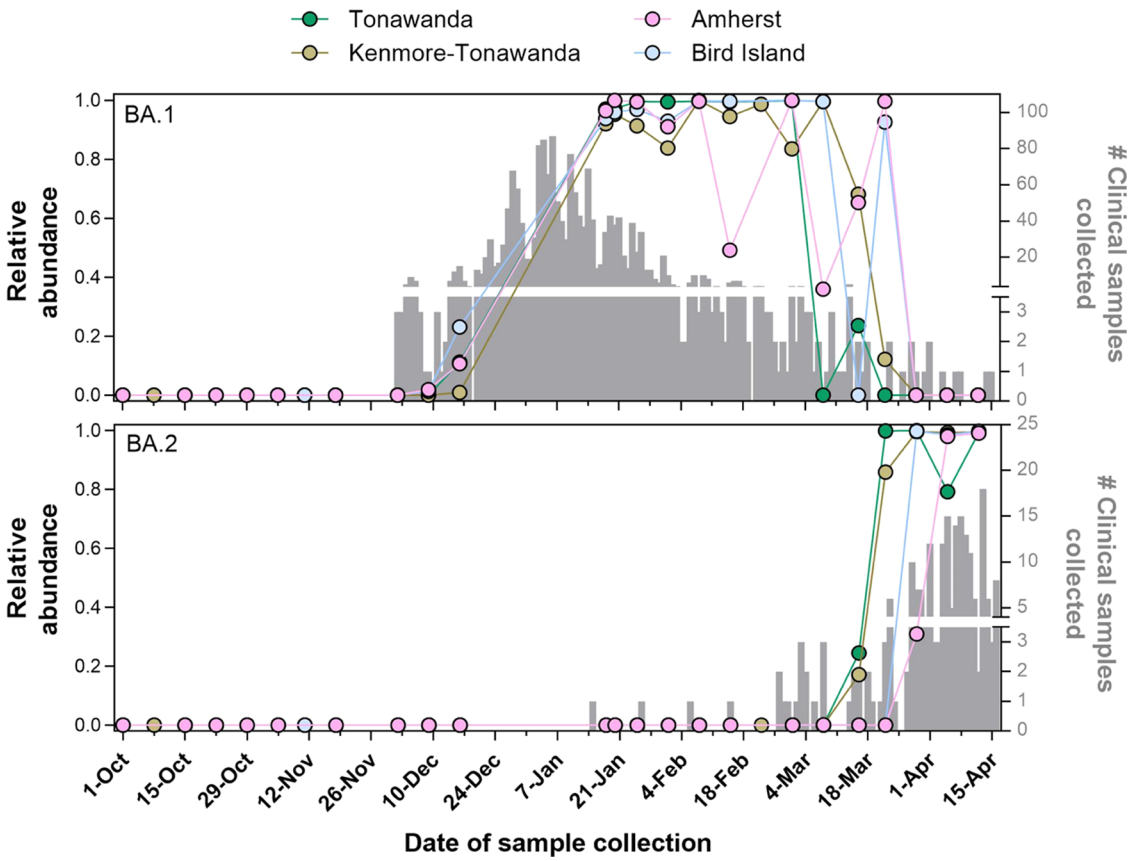


Figure 3. Detection of Omicron BA.1 and BA.2 in wastewater and clinical samples. The relative abundance of Omicron was predicted by Freyja pipeline with the wastewater sequencing data filtered at 50× depth. The relative abundance of BA.1 (top) is the sum of relative abundances of B.1.1.529, BA.1, BA.1.1, and other BA.1.*. The relative abundance of BA.2 (bottom) comprises BA.2.12, BA.2.12.1, and other BA.2.*. The number of clinical samples collected for sequencing was counted on every collection date.

Amherst, and 65.2% (15/23) for Bird Island (Table S4). Although the Bird Island sewershed serves a population that is larger than the other three (Table S1), similar average C_T values were obtained for the SARS-CoV-2 N gene (33.5 for Tonawanda, 33.4 for Kenmore−Tonawanda, 34.2 for Amherst, and 34.1 for Bird Island) (Table S4), indicating that similar genome levels were available for sequencing. Notably, the Bird Island wastewater treatment plant collects wastewater exclusively through combined sewer systems (Table S1). A large fraction of sequencing inhibitors, such as humic acids and heavy metals, in the stormwater runoff may slightly reduce the genome sequencing success rate.[45,46]

Influence of COVID-19 Incidence. The probability of sequencing success was >0.9 when the COVID-19 incidence rate was >33/100000 persons but decreased to 0.75 when the incidence rate was 25.2/100 000 persons (Figure 1B). A previous study reported a probability of 0.75 to quantify SARS-CoV-2 gene in wastewater by RT-qPCR when the 14-day COVID-19 case rate was 152/100 000 persons.[47] In that study, viruses were concentrated from solids-removed influent samples by 10 kDa Centricon ultrafilters.[47] We are not aware of other wastewater sequencing studies that have calculated SARS-CoV-2 genome sequencing success rates with regard to the incidence rate, but our results suggest that SARS-CoV-2 genome sequencing with the viruses recovered by tangential-flow filtration method can be successfully applied at low COVID-19 incidence rates.

Lineage Distributions Estimated by Wastewater and Clinical Data. Sequencing depths play a role in detecting low-abundance mutations, yet there is no depth cutoff recommendation for analyzing wastewater genome sequencing data. Low-depth cutoff (e.g., 10×) allows the use of more genetic information for lineage estimates but may increase the probability of identification errors. In our results, the estimated lineage abundances at sequencing depths of 10× and 50× were very similar (Figure S8). We reported lineage distributions estimated by the genome sequencing data filtered at a depth of 50× to better control the sequencing quality.

Nine dominant groups of lineages were identified in the wastewater samples: AY.* (Delta), B.1.617.2 (Delta), B.1.2, B.1.1.529 (Omicron), BA.1/BA.1.1 (Omicron), BA.1.1.16 (Omicron), other BA.1.* (Omicron), BA.2.12/BA.2.12.1 (Omicron), and other BA.2.* (Omicron). The prevalence of Delta AY.* lineages in wastewater in October−December 2021 was coincident with the lineages observed via clinical surveillance (Figure 2). Two Omicron infection waves in 2022 (BA.1 in January/February and BA.2 in March/April) predicted from the wastewater data aligned with clinical data (Figure 2). Remarkably, wastewater sequencing revealed lineages that were underrepresented or not present in the clinical data. Specifically, the B.1.2 lineage was mainly identified in the wastewater samples from Kenmore−Tonawanda sewershed in October 2021 to early February 2022 (Figure 2A). Moreover, in January/February of 2022, the BA.1/BA.1.1 lineages were prevalent in patient samples

https://doi.org/10.1021/acsestwater.2c00456
ACS EST Water 2023, 3, 954−962

(Figure 2A), whereas the BA.1.1.16 lineage dominated in the wastewater samples (Figure 2B−E). Notably, the genome sequences of the BA.1/BA.1.1 and BA.1.1.16 lineages are similar, which might affect the predictions of their relative abundance.

The reason that B.1.2 lineage was identified in wastewater samples but largely absent from patient samples is unclear. We further investigated the B.1.2 lineage identified in the wastewater samples from Kenmore−Tonawanda. The B.1.2 lineage occurred in patient samples in the area between September 2020 through February 2021 but has remained largely absent from patient samples since then. The Freyja pipeline identified B.1.2 on the basis of a mutation spectrum consisting of 13 nucleotide substitutions (A18424G, A23403G, C10319T, C1059T, C14408T, C21304T, C241T, C27964T, C28472T, C28869T, C3037T, G25563T, and G25907T), three of which (C1059T, C21304T, and G25563T) overlap the mutations in highly abundant Delta lineages circulating at the time, and the other 10 mutations are specific to B.1.2 lineage. Regardless of sharing the three identifying mutations with the Delta lineages in Freyja, the B.1.2 mutation spectrum is not strongly correlated with any of the Delta sublineages (Figure S9). Mining through the patient samples sequenced at the time in Erie County showed that the majority of the B.1.2-specific mutations were rarely identified, including A18424G, C10319T, C27964T, C28472T, C28869T, and G25907T. Our wastewater samples with B.1.2 identified contained most of the B.1.2-specific mutations, except for C1059T, and those regions were sequenced at a depth >100× (Figure S7). Taken together, our results suggest that certain mutation combinations in wastewater sequencing data may cause deconvolution algorithms to bin samples into lineages that are not detected in patient WGS data. It is possible that these lineages remained within the gastrointestinal tract of individuals, possibly immunocompromised individuals with long-term infections.[48−50]

Early Detection of Variants in Wastewater. The values of early detection of SARS-CoV-2 variants in wastewater depend on the delays in clinical analysis.[51] The workflow optimization for a quick turnaround of genome sequencing analysis is beyond the scope of this study. Here, we only compared the detection of emerging SARS-CoV-2 variants based on the date of collection for both wastewater and patient samples.

The Omicron BA.1 was first identified in Amherst and Bird Island wastewater with a relative abundance of 1−2% in the second week of December, which is ∼1 week later than the patient-derived samples (Figure 3). The detection of Omicron BA.2 from wastewater samples in March was similarly delayed (Figure 3). Note that some BA.2 patient-derived samples were collected in January and February, but there was no Omicron BA.2 outbreak in the studied sewersheds until March. The poor detection of Omicron BA.2 in wastewater may be attributable to the relatively low sequence coverage of the SARS-CoV-2 genome from wastewater during mid-February and early March (<65%; Table S3). Nevertheless, our findings suggest that wastewater is a relevant proxy for patient samples. Wastewater genome sequencing should be considered in the face of practical delays in clinical sequencing for early detection of variants.

The sensitivity of variant detection was also compared between RT-qPCR and wastewater sequencing. The RT-qPCR assays were conducted to target BA.1-associated mutations delH69/V70 and BA.2-associated mutations delL24/P25/P26 and A27S in the S gene. Given that delH69/V70 was also seen in the Alpha variant, we detected additional Omicron-associated mutations Q493R and Q498R within the S gene. In the wastewater sequencing data, we searched for the presence of all of the above mutations. Note that the regions where the mutations are located (nucleotides 21 500−22 500 and 23 000−23 250) were all well sequenced with depths >100× in most of the samples (Figure S7). As a result, we found that RT-qPCR assays were more sensitive than wastewater genome sequencing for detecting Omicron BA.1-specific mutations, but wastewater genome sequencing was more sensitive for Omicron BA.2-specific mutations (Table S4). However, RT-qPCR assays were reported less sensitive than digital PCR assays,[52] which could detect variant-specific mutations earlier than genetic sequencing.[9]

Interestingly, our RT-qPCR assays detected delH69/V70 mutations but not Q493R and Q498R in the Kenmore−Tonawanda wastewater samples collected in October−November 2021 (Table S4), suggesting that the positive detection of delH69/V70 is not associated with Omicron variants. These RT-qPCR results were further confirmed with the wastewater sequencing data (Table S4). When we searched the patient WGS data collected at the time, we did not identify patient-derived lineages that contain delH69/V70. It is possible that certain unknown lineages present in wastewater contribute to the detection of delH69/V70 mutations.

CONCLUSIONS

Our research demonstrates that tangential-flow filtration to concentrate viruses from wastewater samples enables the extraction of nucleic acids of high-enough quality for stable performance of SARS-CoV-2 genome sequencing. Complete or near-complete genomes at a depth of 10× were sequenced from 68% (64/94) of the wastewater samples. Moreover, the tangential-flow filtration method demonstrates the successful and reliable sequencing of SARS-CoV-2 genomes at lower COVID-19 incidence rates. Our results report a 0.9 probability of genome sequencing success when the COVID-19 incidence rate exceeded 33/100 000 people. Furthermore, genome sequencing of SARS-CoV-2 in wastewater revealed lineages underrepresented or not detected from patient samples. Future studies are needed to advance wastewater genome sequencing data interpretation and streamline the workflow to shorten the turnaround time for early detection.

The developed reusable and likely cost-effective tangential-flow filtration method is readily applicable to other sequencing assays that require high-quality viral nucleic acids from wastewater. Given the flexibility of concentrating large volumes of wastewater, the tangential-flow filtration system can be further optimized for genomic surveillance of any low-abundance viruses in wastewater. To implement the practicality of the tangential-flow filtration system, we recommend incorporating an automated control system into membrane filtration and washing steps to scale up for high-throughput sample processing at the same time. Future work is also needed to monitor and extend the lifetime of ultrafilter membranes in long-term applications to save costs.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsestwater.2c00456.

https://doi.org/10.1021/acsestwater.2c00456
ACS EST Water 2023, 3, 954−962

Inhibition test; RT-qPCR assay details and limit of detection (LOD) and limit of quantification (LOQ) determination; details of four sewersheds in this study; details of RT-qPCR assay for quantifying SARS-CoV-2 N gene and S gene mutations; checklist of RT-qPCR experiments according to the MIQE (minimum information for publication of quantitative real-time PCR experiments) guidelines; summary of wastewater samples sequenced in this study; geographic locations of four sewersheds in Erie County, New York; schematic of tangential-flow filtration system; SARS-CoV-2 N gene levels in wastewater and blank test samples; determination of limit of detection (LOD; left) and limit of quantification (LOQ; right) of SARS-CoV-2 N gene by RT-qPCR; RT-qPCR inhibition test of SARS-CoV-2 N gene in the wastewater nucleic acid extracts; comparison of SARS-CoV-2 genome coverage from wastewater with different workflows; heat map of read depths of SARS-CoV-2 whole-genome sequencing from wastewater samples; comparison of relative abundances of SARS-CoV-2 lineages estimated by the sequencing data at depths of 10× and 50×; pairwise Pearson correlation of B.1.2 lineage mutation spectrum against all other lineage mutation spectra (PDF)

AUTHOR INFORMATION

Corresponding Author

Yinyin Ye − *Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, New York 14260, United States;* orcid.org/0000-0003-0625-9657; Phone: (716) 645-4002; Email: yinyinye@buffalo.edu

Authors

Emily Segelhurst − *Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, New York 14260, United States;* orcid.org/0000-0001-5786-2988

Jonathan E. Bard − *UB Genomics and Bioinformatics Core, Department of Biochemistry, Jacobs School of Medicine and Biomedical Sciences, and Genetics, Genomics and Bioinformatics Graduate Program, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, New York 14203, United States*

Annemarie N. Pillsbury − *Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, New York 14260, United States*

Md Mahbubul Alam − *Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, New York 14260, United States*

Natalie A. Lamb − *UB Genomics and Bioinformatics Core, University at Buffalo, Buffalo, New York 14203, United States*

Chonglin Zhu − *Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, New York 14260, United States*

Alyssa Pohlman − *UB Genomics and Bioinformatics Core, University at Buffalo, Buffalo, New York 14203, United States*

Amanda Boccolucci − *UB Genomics and Bioinformatics Core, University at Buffalo, Buffalo, New York 14203, United States*

Jamaal Emerson − *Department of Microbiology and Immunology, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, New York 14203, United States*

Brandon J. Marzullo − *UB Genomics and Bioinformatics Core and Department of Biochemistry, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, New York 14203, United States;* orcid.org/0000-0002-3547-1182

Donald A. Yergeau − *UB Genomics and Bioinformatics Core, University at Buffalo, Buffalo, New York 14203, United States*

Norma J. Nowak − *UB Genomics and Bioinformatics Core and Department of Biochemistry, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, New York 14203, United States*

Ian M. Bradley − *Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, New York 14260, United States; Research and Education in Energy, Environmental and Water (RENEW) Institute, University at Buffalo, Buffalo, New York 14260, United States;* orcid.org/0000-0001-5338-1167

Jennifer A. Surtees − *Department of Biochemistry, Jacobs School of Medicine and Biomedical Sciences, Department of Microbiology and Immunology, Jacobs School of Medicine and Biomedical Sciences, and Genetics, Genomics and Bioinformatics Graduate Program, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, New York 14203, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsestwater.2c00456

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by Erie County Department of Health (ECDOH) award 91287 to Y.Y. and I.M.B. and award 91279 to J.A.S. The authors thank undergraduate students Vicky Huang and Danya Hanin for processing the wastewater samples in the lab. The authors also thank Joseph L. Fiegl for coordinating the wastewater sample collection. The GISAID data used in the manuscript (Identifier: EPI_SET_220906wr) is composed of 4952 individual genome sequences published in GISAID's EpiCoV database. To view the contributors of each individual sequence with details, visit doi.org/10.55876/gis8.220906wr. Most of the samples (4653/4952) were sequenced at the University at Buffalo. The authors are grateful for the sequencing work of scientists and healthcare professionals who generated the remaining sequences from the region and made them publicly accessible through GISAID.

REFERENCES

(1) Walensky, R. P.; Walke, H. T.; Fauci, A. S. SARS-CoV-2 variants of concern in the United States-challenges and opportunities. *JAMA* **2021**, *325*, 1037−1038.

(2) Munnink, B. B. O.; Nieuwenhuijse, D. F.; Stein, M.; O'Toole, A.; Haverkate, M.; Mollers, M.; Kamga, S. K.; Schapendonk, C.; Pronk, M.; Lexmond, P.; et al. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat. Med.* **2020**, *26*, 1405−1410.

(3) Crits-Christoph, A.; Kantor, R. S.; Olm, M. R.; Whitney, O. N.; Al-Shayeb, B.; Lou, Y. C.; Flamholz, A.; Kennedy, L. C.; Greenwald, H.; Hinkle, A.; et al. Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants. *mBio* **2021**, *12*, No. e02703-20.
(4) Nemudryi, A.; Nemudraia, A.; Wiegand, T.; Surya, K.; Buyukyoruk, M.; Cicha, C.; Vanderwood, K. K.; Wilkinson, R.; Wiedenheft, B. Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater. *Cell Rep. Med.* **2020**, *1*, No. 100098.
(5) Karthikeyan, S.; Levy, J. I.; De Hoff, P.; Humphrey, G.; Birmingham, A.; Jepsen, K.; Farmer, S.; Tubb, H. M.; Valles, T.; Tribelhorn, C. E.; et al. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature* **2022**, *609*, 101−108.
(6) Smyth, D. S.; Trujillo, M.; Gregory, D. A.; Cheung, K.; Gao, A.; Graham, M.; Guan, Y.; Guldenpfennig, C.; Hoxie, I.; Kannoly, S.; et al. Tracking cryptic SARS-CoV-2 lineages detected in NYC wastewater. *Nat. Commun.* **2022**, *13*, No. 635.
(7) Sims, D.; Sudbery, I.; Ilott, N. E.; Heger, A.; Ponting, C. P. Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* **2014**, *15*, 121−132.
(8) Valieris, R.; Drummond, R. D.; Defelicibus, A.; Dias-Neto, E.; Rosales, R. A.; Tojal da Silva, I. A mixture model for determining SARS-Cov-2 variant composition in pooled samples. *Bioinformatics* **2022**, *38*, 1809−1815.
(9) Lou, E. G.; Sapoval, N.; McCall, C.; Bauhs, L.; Carlson-Stadler, R.; Kalvapalle, P.; Lai, Y.; Palmer, K.; Penn, R.; Rich, W.; et al. Direct comparison of RT-ddPCR and targeted amplicon sequencing for SARS-CoV-2 mutation monitoring in wastewater. *Sci. Total Environ.* **2022**, *833*, No. 155059.
(10) Graham, K. E.; Loeb, S. K.; Wolfe, M. K.; Catoe, D.; Sinnott-Armstrong, N.; Kim, S.; Yamahara, K. M.; Sassoubre, L. M.; Grijalva, L. M. M.; Roldan-Hernandez, L.; et al. SARS-CoV-2 RNA in wastewater settled solids is associated with COVID-19 cases in a large urban sewershed. *Environ. Sci. Technol.* **2021**, *55*, 488−498.
(11) LaTurner, Z. W.; Zong, D. M.; Kalvapalle, P.; Gamas, K. R.; Terwilliger, A.; Crosby, T.; Ali, P.; Avadhanula, V.; Santos, H. H.; Weesner, K.; et al. Evaluating recovery, cost, and throughput of different concentration methods for SARS-CoV-2 wastewater-based epidemiology. *Water Res.* **2021**, *197*, No. 117043.
(12) Zheng, X.; Deng, Y.; Xu, X.; Li, S.; Zhang, Y.; Ding, J.; On, H. Y.; Lai, J. C. C.; In Yau, C.; Chin, A. W. H.; et al. Comparison of virus concentration methods and RNA extraction methods for SARS-CoV-2 wastewater surveillance. *Sci. Total Environ.* **2022**, *824*, No. 153687.
(13) Peccia, J.; Zulli, A.; Brackney, D. E.; Grubaugh, N. D.; Kaplan, E. H.; Casanovas-Massana, A.; Ko, A. I.; Malik, A. A.; Wang, D.; Wang, M.; et al. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nat. Biotechnol.* **2020**, *38*, 1164−1167.
(14) Pérez-Cataluña, A.; Chiner-Oms, A.; Cuevas-Ferrando, E.; Diaz-Reolid, A.; Falco, I.; Randazzo, W.; Giron-Guzman, I.; Allende, A.; Bracho, M. A.; Comas, I.; Sánchez, G. Spatial and temporal distribution of SARS-CoV-2 diversity circulating in wastewater. *Water Res.* **2022**, *211*, No. 118007.
(15) Fontenele, R. S.; Kraberger, S.; Hadfield, J.; Driver, E. M.; Bowes, D.; Holland, L. A.; Faleye, T. O. C.; Adhikari, S.; Kumar, R.; Inchausti, R.; et al. High-throughput sequencing of SARS-CoV-2 in wastewater provides insights into circulating variants. *Water Res.* **2021**, *205*, No. 117710.
(16) Lin, X.; Glier, M.; Kuchinski, K.; Ross-Van Mierlo, T.; McVea, D.; Tyson, J. R.; Prystajecky, N.; Ziels, R. M. Assessing Multiplex Tiling PCR Sequencing Approaches for Detecting Genomic Variants of SARS-CoV-2 in Municipal Wastewater. *mSystems* **2021**, *6*, No. e01068-21.
(17) Bar-Or, I.; Weil, M.; Indenbaum, V.; Bucris, E.; Bar-Ilan, D.; Elul, M.; Levi, N.; Aguvaev, I.; Cohen, Z.; Shirazi, R.; et al. Detection of SARS-CoV-2 variants by genomic analysis of wastewater samples in Israel. *Sci. Total Environ.* **2021**, *789*, No. 148002.
(18) Izquierdo-Lara, R.; Elsinga, G.; Heijnen, L.; Munnink, B. B. O.; Schapendonk, C. M. E.; Nieuwenhuijse, D.; Kon, M.; Lu, L.; Aarestrup, F. M.; Lycett, S.; et al. Monitoring SARS-CoV-2 circulation and diversity through community wastewater sequencing, the Netherlands and Belgium. *Emerg. Infect. Dis.* **2021**, *27*, 1405−1415.
(19) Furtak, V.; Roivainen, M.; Mirochnichenko, O.; Zagorodnyaya, T.; Laassri, M.; Zaidic, S. Z.; Rehman, L.; Alam, M. M.; Chizhikov, V.; Chumakov, K. Environmental surveillance of viruses by tangential flow filtration and metagenomic reconstruction. *Eurosurveillance* **2016**, *21*, No. e15-30193.
(20) Bisseux, M.; Colombet, J.; Mirand, A.; Roque-Afonso, A. M.; Abravanel, F.; Izopet, J.; Archimbaud, C.; Peigue-Lafeuille, H.; Debroas, D.; Bailly, J. L.; Henquell, C. Monitoring human enteric viruses in wastewater and relevance to infections encountered in the clinical setting: a one-year experiment in central France, 2014 to 2015. *Eurosurveillance* **2018**, *23*, No. e17-00237.
(21) Langenfeld, K.; Chin, K.; Roy, A.; Wigginton, K.; Duhaime, M. B. Comparison of ultrafiltration and iron chloride flocculation in the preparation of aquatic viromes from contrasting sample types. *PeerJ* **2021**, *9*, No. e11111.
(22) Thurber, R. V.; Haynes, M.; Breitbart, M.; Wegley, L.; Rohwer, F. Laboratory procedures to generate viral metagenomes. *Nat. Protoc.* **2009**, *4*, 470−483.
(23) Shi, X.; Tal, G.; Hankins, N. P.; Gitis, V. Fouling and cleaning of ultrafiltration membranes: A review. *J. Water Process Eng.* **2014**, *1*, 121−138.
(24) Lu, X.; Wang, L.; Sakthivel, S. K.; Whitaker, B.; Murray, J.; Kamili, S.; Lynch, B.; Malapati, L.; Burke, S. A.; Harcourt, J.; et al. US CDC real-time reverse transcription PCR panel for detection of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerg. Infect. Dis.* **2020**, *26*, 1654.
(25) Lee, W. L.; Gu, X.; Armas, F.; Wu, F.; Chandra, F.; Chen, H.; Xiao, A.; Leifels, M.; Chua, F. J. D.; Kwok, G. W. C.et al.Quantitative detection of SARS-CoV-2 Omicron BA.1 and BA.2 variants in wastewater through allele-specific RT-qPCR, 2022. DOI: 10.1101/2021.12.21.21268077.
(26) Lee, W. L.; Imakaev, M.; Armas, F.; McElroy, K. A.; Gu, X. Q.; Duvallet, C.; Chandra, F.; Chen, H. J.; Leifels, M.; Mendola, S.; et al. Quantitative SARS-CoV-2 Alpha variant B.1.1.7 tracking in wastewater by allele-specific RT-qPCR. *Environ. Sci. Technol. Lett.* **2021**, *8*, 675−682.
(27) He, H.; Zhou, P.; Shimabuku, K. K.; Fang, X.; Li, S.; Lee, Y.; Dodd, M. C. Degradation and deactivation of bacterial antibiotic resistance genes during exposure to free chlorine, monochloramine, chlorine dioxide, ozone, ultraviolet light, and hydroxyl radical. *Environ. Sci. Technol.* **2019**, *53*, 2013−2026.
(28) Bustin, S. A.; Benes, V.; Garson, J. A.; Hellemans, J.; Huggett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffl, M. W.; Shipley, G. L.; et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **2009**, *55*, 611−622.
(29) nCoV-2019 Sequencing Protocol V.1. https://www.protocols.io/view/ncov-2019-sequencing-protocol-bp2l6n26rgqe/v1?version_warning=no (accessed January 22, 2020).
(30) Wingett, S. W.; Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* **2018**, *7*, 1338.
(31) Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754−1760.
(32) Danecek, P.; Bonfield, J. K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M. O.; Whitwham, A.; Keane, T.; McCarthy, S. A.; Davies, R. M.; Li, H. Twelve years of SAMtools and BCFtools. *GigaScience* **2021**, *10*, No. giab008.
(33) Erie County SARS-CoV-2 Wastewater Monitoring Dashboard. https://erieny.maps.arcgis.com/apps/dashboards/a95853269eec489ea59e5b71571f2e76 (accessed January 12, 2021).
(34) Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges* **2017**, *1*, 33−46.

(35) Lamb, N. A.; Bard, J. E.; Pohlman, A.; Boccolucci, A.; Yergeau, D. A.; Marzullo, B. J.; Pope, C.; Burstein, G.; Tomaszewski, J.; Nowak, N. J.; Surtees, J. A. Genomic Surveillance of SARS-CoV-2 in Erie County, New York, 2021. DOI: 10.1101/2021.07.01.21259869.
(36) Bewick, V.; Cheek, L.; Ball, J. Statistics review 14: Logistic regression. *Crit. Care* **2005**, *9*, 112−118.
(37) Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer-Verlag: New York, 2016.
(38) *Integrated Development for RStudio*; PBC: Boston, MA, 2020.
(39) Jahn, K.; Dreifuss, D.; Topolsky, I.; Kull, A.; Ganesanandamoorthy, P.; Fernandez-Cassi, X.; Bänziger, C.; Devaux, A. J.; Stachler, E.; Caduff, L.; et al. Early detection and surveillance of SARS-CoV-2 genomic variants in wastewater using COJAC. *Nat. Microbiol.* **2022**, *7*, 1151−1160.
(40) Lambisia, A. W.; Mohammed, K. S.; Makori, T. O.; Ndwiga, L.; Mburu, M. W.; Morobe, J. M.; Moraa, E. O.; Musyoki, J.; Murunga, N.; Mwangi, J. N.; et al. Optimization of the SARS-CoV-2 ARTIC Network V4 Primers and Whole Genome Sequencing Protocol. *Front. Med.* **2022**, *9*, No. 836728.
(41) Clark, C. R.; Hardison, M. T.; Houdeshell, H. N.; Vest, A. C.; Whitlock, D. A.; Skola, D. D.; Koble, J. S.; Oberholzer, M.; Schroth, G. P. Evaluation of an optimized protocol and Illumina ARTIC V4 primer pool for sequencing of SARS-CoV-2 using COVIDSeq and DRAGEN COVID Lineage App workflow, 2022. DOI: 10.1101/2022.01.07.475443.
(42) Mantilla-Calderon, D.; Huang, K. Y.; Li, A. J.; Chibwe, K.; Yu, X. Q.; Ye, Y.; Liu, L.; Ling, F. Q. Emerging investigator series: meta-analyses on SARS-CoV-2 viral RNA levels in wastewater and their correlations to epidemiological indicators. *Environ. Sci.: Water. Res. Technol.* **2022**, *8*, 1391−1407.
(43) Itokawa, K.; Sekizuka, T.; Hashino, M.; Tanaka, R.; Kuroda, M. Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. *PLoS One* **2020**, *15*, No. e0239403.
(44) Agrawal, S.; Orschler, L.; Lackner, S. Metatranscriptomic analysis reveals SARS-CoV-2 mutations in wastewater of the Frankfurt metropolitan area in southern Germany. *Microbiol. Resour. Announc.* **2021**, *10*, No. e00280-21.
(45) Grant, S. B.; Rekhi, N. V.; Pise, N. R.; Reeves, R. L.; Matsumoto, M.; Wistrom, A.; Moussa, L.; Bay, S.; Kayhanian, M. A review of the contaminants and toxicity associated with particles in stormwater runoff. *Terminology* **2003**, *2*, 1−173.
(46) Schrader, C.; Schielke, A.; Ellerbroek, L.; Johne, R. PCR inhibitors - occurrence, properties and removal. *J. Appl. Microbiol.* **2012**, *113*, 1014−1026.
(47) Tiwari, A.; Lipponen, A.; Hokajarvi, A. M.; Luomala, O.; Sarekoski, A.; Rytkonen, A.; Osterlund, P.; Al-Hello, H.; Juutinen, A.; Miettinen, I. T.; et al. Detection and quantification of SARS-CoV-2 RNA in wastewater influent in relation to reported COVID-19 incidence in Finland. *Water Res.* **2022**, *215*, No. 118220.
(48) Shafer, M. M.; Gregory, D.; Bobholz, M. J.; Roguet, A.; Haddock Soto, L. A.; Rushford, C.; Vuyk, W. C.; Janssen, K. H.; Ries, H. J.; Pilch, H. E.et al. Tracing the origin of SARS-CoV-2 Omicron-like Spike sequences detected in wastewater, 2022. DOI: 10.1101/2022.10.28.22281553.
(49) Wilkinson, S. A. J.; Richter, A.; Casey, A.; Osman, H.; Mirza, J. D.; Stockton, J.; Quick, J.; Ratcliffe, L.; Sparks, N.; Cumley, N.; et al. Recurrent SARS-CoV-2 mutations in immunodeficient patients. *Virus Evol.* **2022**, *8*, No. veac050.
(50) Corey, L.; Beyrer, C.; Cohen, M. S.; Michael, N. L.; Bedford, T.; Rolland, M. SARS-CoV-2 Variants in Patients with Immunosuppression. *N. Engl. J. Med.* **2021**, *385*, 562−566.
(51) Larsen, D. A.; Wigginton, K. R. Tracking COVID-19 with wastewater. *Nat. Biotechnol.* **2020**, *38*, 1151−1153.
(52) Ahmed, W.; Smith, W. J.; Metcalfe, S.; Jackson, G.; Choi, P. M.; Morrison, M.; Field, D.; Gyawali, P.; Bivins, A.; Bibby, K.; Simpson, S. L. Comparison of RT-qPCR and RT-dPCR platforms for the trace detection of SARS-CoV-2 RNA in wastewater. *ACS ES&T Water* **2022**, *2*, 1871−1880.