Environmental Science
Water Research & Technology

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Environ. Sci.: Water Res. Technol.*, 2025, **11**, 1494

Dynamics of SARS-CoV-2 variants in southwest Ohio municipal wastewater†

Maitreyi Nagarkar, ‡[a] Scott P. Keely,[a] Emily A. Wheaton,[a] Chloe Hart, §[a] Michael A. Jahne,[a] Jay L. Garland,[a] Eunice Varughese[ab] and Nichole E. Brinkman *[a]

Received 19th February 2025,
Accepted 8th April 2025

DOI: 10.1039/d5ew00169b

rsc.li/es-water

Wastewater surveillance has proven to be a widely useful means for tracking the dynamics of COVID-19, particularly as the emphasis on clinical testing and reporting of case data continues to decline. Here we present wastewater monitoring data from a multi-year sampling campaign at 11 wastewater collection facilities in Ohio. We found strong correlations between flow-adjusted wastewater concentrations of the virus (as represented through quantification of N2 gene fragments) and reported cases and used sequencing to confirm the sequential arrival of several variants of concern (VOCs) between winter 2020 and spring 2022. We observed that the three main VOCs in our dataset, alpha, delta, and omicron, showed differing temporal dynamics like length of time from first detection to dominating the wastewater signal. We also found credible variation in the relationship between wastewater concentration and clinical cases during different periods within our time series (delineated based on the dominant VOC), indicating the possibility of differential fecal shedding by the three variants.

Water impact

Wastewater surveillance has been established as a powerful tool to track the infection dynamics of COVID-19. In southwest Ohio sewersheds, the early SARS-CoV-2 variants, alpha, delta and omicron, showed varied temporal dynamics and relationships with clinical cases, likely due to differential fecal shedding rates exhibited by the variants.

Introduction

Wastewater monitoring for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA has become a globally adopted means of tracking both quantitative trends and taxonomic diversity of the virus since the inception of the COVID-19 pandemic. Sampling of wastewater to detect and quantify SARS-CoV-2, has been successfully conducted from samples with a wide variety of input populations, from individual buildings like college dormitories[1–3] to wastewater treatment plants serving hundreds of thousands of people,[4–6] to bodies of water receiving treated or untreated effluent.[7,8] Furthermore, it is now well-established that wastewater trends are closely associated with case counts and other disease surveillance indicators,[6,9–11] despite large variations in sample collection and processing methodologies, sewershed size, matrix composition, and more. This robustness supports the utility of wastewater measurements as a reliable indicator of community disease regardless of the level of testing and reporting occurring in a community.

The state of Ohio was an early adopter of wastewater surveillance, forming the Ohio Coronavirus Wastewater Monitoring Network (OCWMN) in May 2020 as tasked by the governor.[12] This partnership of the state health department, academic labs, non-governmental institutions, municipal sewer districts, the Ohio Environmental Protection Agency (EPA), and the US EPA Office of Research and Development led to twice weekly sampling at over 70 sites throughout the state, capturing more than half of Ohio's population. Multiple analyses of the resulting datasets have indicated flow-adjusted SARS-CoV-2 N2 gene marker concentrations are strongly correlated between wastewater data and case counts.[9,13] Results continue to be conveyed to the public through a dashboard available on the Ohio Department of Health's website, which displays a map of

[a] *U.S. Environmental Protection Agency, Office of Research and Development, 26 W. Martin Luther King Drive, Cincinnati, OH 45268, USA.*
E-mail: brinkman.nichole@epa.gov
[b] *U.S. Environmental Protection Agency, Region 6, 1201 Elm Street, Dallas, TX 75270, USA*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d5ew00169b

‡ Currently, The Water Research Foundation, 1199 North Fairfax Street, Alexandria, VA 22314, USA.
§ Currently, Battelle Memorial Institute, 505 King Ave, Columbus, OH 43201, USA.

sampling sites along with daily N2 gene copy loads and an indicator of current wastewater trends at each of them. Sampling has continued through at least winter 2024, with additional viral targets (respiratory syncytial virus, influenza A and B) now included https://data.ohio.gov/wps/portal/gov/data/view/covid-19-reporting).

At many study sites globally, initial wastewater surveillance efforts focused on detection and quantification of viral RNA were quickly followed by sequencing endeavors, which recover genetic mutations present in wastewater to characterize the diversity of circulating SARS-CoV-2 viruses. As the virus quickly evolved to new, potentially more transmissible or virulent variants, the US Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) identified some of these as variants of concern (VOCs). In Ohio, wastewater sample sequencing was initiated in April 2021 and allele frequencies at genomic loci associated with VOCs were reported to the Ohio Department of Health on a regular basis. The utility of wastewater sequencing has also become well-established: multiple studies have shown good correspondence between wastewater sequencing and clinical sequencing data[14,15] and demonstrated that sequencing holds promise as an early detection tool for new variants.[5,16,17] Additionally, coupling knowledge of the concentration of viral signal with the phylogenetic makeup may allow researchers to better understand how different variants spread and evolve over time. As the frequency and reliability of clinical testing and reporting continues to decrease for COVID-19, wastewater monitoring continues to offer consistent data for public health.

In support of the OCWMN, we processed and analyzed samples from October 2020 through March 2022 at 11 sampling sites in Southwestern Ohio. This set of samples captured the initial winter wave as well as the subsequent waves of the alpha, delta, and omicron variants. We report trends in raw and flow-adjusted wastewater measurements and model the relationship between wastewater measurements and reported cases, both quantitatively and from sequencing data, during the different phases of the pandemic (when distinct variants of concern were predominant). We hypothesized that COVID-19 cases can be explained by the interaction of the SARS-CoV-2 loads in wastewater (generated by ddPCR) and phase of the pandemic (which variant is predominant as determined from sequencing data). As wastewater surveillance continues to be implemented for monitoring both SARS-CoV-2 and other emerging pathogens, this work demonstrates how retrospective analysis can be used to refine our understanding of the relationship between wastewater data and disease prevalence.

Methods

Sample collection

24 h composite wastewater samples were collected from 11 wastewater treatment plants (WWTPs), water reclamation facilities (WRFs), and water pollution control plants (WPCs) in Southwestern Ohio (Fig. 1, Table 1, labeled sewershed (SS) 1–11).

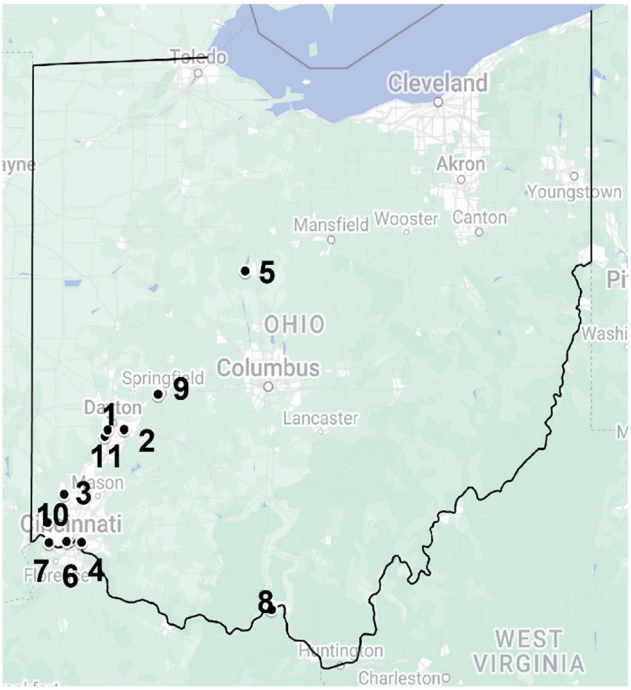

Fig. 1 Map of wastewater treatment plant sampling locations.

On the day of completion of the 24 h composite sampling, 1 L of flow-weighted composited sample was transferred on ice to the US EPA AWBERC facility in autoclaved polypropylene screw-cap bottles. Upon receipt to the laboratory, bottles were disinfected with 70% ethyl alcohol and stored at 4 °C until processing.

Sample concentration

The composite samples were homogenized by inverting; then aliquoted into two 225 mL subsamples. Each of these replicates was amended with 10× RNase-free phosphate-buffered saline (PBS; ThermoFisher Scientific, Waltham MA) to a 1× PBS concentration. Each bottle was also spiked with

Table 1 Sewershed characteristics

Facility code	Sewershed type	Average flow rate (mgd)	Population served	Industrial flow (%)
SS1	Separate	45	226 729	18.8
SS2	Separate	9.1	75 509	2.6
SS3	Separate	8	65 000	0.375
SS4	Combined	31	200 666	4.8
SS5	Combined	10.5	36 000	9.5
SS6	Combined	115	566 136	7.8
SS7	Combined	13	125 247	38.4
SS8	Combined	2	20 366	0
SS9	Combined	9.57	88 710	6.9
SS10	Separate	3	83 058	0
SS11	Separate	15.5	113 629	3.7

approximately 10^7 viral particles of *Betacoronovirus-1* strain OC43 (ATCC, Manassas, VA) as an exogenous internal processing control. Samples were concentrated as described previously.[13] Briefly, each subsample was centrifuged at 3000 × *g* for 15 minutes to collect the pelleted material for nucleic acid extraction. The supernatant was then filtered through a 0.45 μm mixed cellulose ester membrane filter. Material collected on the membrane filter was also subject to extraction. Both pellet and filter fractions were stored at −20 °C until extraction.

Nucleic acid extraction and RTddPCR

To isolate RNA from the concentrated wastewater samples, cold TRIzol Reagent (1.5 mL, ThermoFisher Scientific) was added to sample pellets or membrane filters in 5 mL PowerWater DNA bead tubes (Qiagen). Samples were vortexed to homogenize *via* bead-beating, and then 0.3 mL chloroform was added. The samples were allowed to partition. Then, RNAse-free glycogen (ThermoFisher Scientific) was added to the aqueous phase and RNA was subsequently precipitated using isopropanol (Fisher Scientific, Waltham MA). Samples were washed with 75% ethanol then allowed to dry before RNA was resuspended in 125 μL DEPC-treated water with 0.1. mM EDTA (ThermoFisher). Filter and pellet RNA extracts were stored separately at −80 °C. Quantities of SARS-CoV-2 N2 gene fragments were determined using the QX200 Droplet Digital PCR System (BioRad), as previously described.[13] Each ddPCR plate included negative ddPCR controls, composed of nuclease-free water, and positive ddPCR controls using commercially available SARS-CoV-2 genomic RNA (ATCC, Manassas, VA). The ddPCR mutation assays were run with the QX200 Droplet Digital PCR System (BioRad) and the One-Step RT-ddPCR Advanced Kit for Probes (BioRad). RT-ddPCR was also used to detect two mutations of the spike gene that are characteristic of the alpha variant. For the N501Y assay, the ddPCR mutation assay: dMDS731762551 (BioRad) was used as instructed by the manufacturer. Thermal cycling conditions were 50 °C for 1 h, then 95 °C for 10 min, followed by 50 cycles of 94 °C for 30 s and 55 °C for 1 min. The enzyme was deactivated at 98 °C for 10 minutes and reactions were held at 4 °C until transfer to the QX200 reader. The 69/70del assay was also run with the QX200 Droplet Digital PCR System and the One-Step RT-ddPCR Advanced Kit for Probes. For this assay, the primers and probes previously described[18] were implemented. Thermal cycling conditions for this assay consisted of 50 °C for 1 h, 95 °C for 10 min, followed by 50 cycles of 94 °C for 30 s and 62 °C for 1 min. The enzyme was deactivated at 98 °C for 10 minutes followed by a 4 °C hold.

Library preparation

RNA obtained from the filter fraction was used for sequencing. RNA extracts from the filter fractions were DNase-treated and reverse-transcribed into cDNA using the SuperScript IV Reverse Transcriptase (ThermoFisher Scientific) according to the manufacturer's instructions. When possible (*i.e.* sufficient yield), extracts were processed both as is and at a 1 : 5 dilution to mitigate impacts of potential inhibition. The SARS-CoV-2 genome was amplified, and amplicons indexed using the IDT xGEN SARS-CoV-2 Amplicon Panel (formerly Swift Biosciences SNAPv2 kit) using the "low viral load input" recommendations for thermocycler programs. Where specified, the xGEN SARS-CoV-2 Sgene (spike) panel was used which consists of all the same reagents and steps, but the primer mix contains only the primers spanning the spike gene region.

Some samples did not move forward to sequencing either because: (1) there was no amplification of SARS-CoV-2 RNA using the N2 assay, (2) the number of N2-positive droplets from ddPCR for the same extract was fewer than 10, or (3) the final concentration of the library (evaluated using KAPA PCR) was less than 0.5 nM.

Sequencing

Each of the libraries was diluted to a concentration of 0.5 nM using 10 mM Tris-HCl, pH 8.5 so that equimolar volumes could be combined in the final library pool. Once libraries were combined, the pool was diluted and denatured according to the Illumina NextSeq Denature and Dilute protocol.[19] The final library, along with 10% PhiX (Illumina) was added to an Illumina NextSeq 300 cycle mid-output or high-output cartridge and sequenced on a NextSeq550.

Bioinformatics

Completed NextSeq runs were demultiplexed on local run manager and data from the four lanes were concatenated using a bash script. We then used fastp to filter sequences based on quality and make contigs, using the default quality thresholds of phred quality ≥15 and a maximum of 40% unqualified bases. Resulting contigs were mapped to the Wuhan SARS-CoV-2 genome (NC_045512) using bwa-mem in Samtools with the default parameters. Primer sequences were removed using the trim command in iVar.[20] Finally, we used LoFreq to make variant calls, including both SNVs and indels.[21]

We also used Freyja,[17] which solves a depth-weighted least absolute deviation regression problem based on SNV frequency across relevant genome sites, to estimate the lineage breakdown in our samples. Freyja uses samtools mpileup and iVar for making SNV calls. We set a minimum depth of 10× for the positions considered. Although the lineage definitions in the Freyja tool include mutations across the entire SARS-CoV-2 genome, we used this tool on both samples where the whole genome was covered and those where only the spike gene was sequenced.

Statistical methods

Evaluation of process control OC43. OC43 was spiked into samples as a process control to monitor performance of sample processing and analysis. The recovery efficiency of the spiked OC43 was assessed from approximately 76% of all

the samples using a Bayesian intercept model that estimates the overall recovery mean (ESI† 1). The mean OC43 recovery efficiency across all sewersheds was 47.4% (95% Bayesian credible interval (BCI): 44.3–50.7%). A comparison of observed recovery data with simulated draws from the Bayesian intercept model showed that the observed values were within the range of the model's 95% BCI. In addition, a *Q*–*Q* plot of the residuals indicated that they are normally distributed. The consistent performance of the OC43 process control across all wastewater samples demonstrates that the laboratory methods are reliable.

SARS-CoV-2 wastewater fecal load. The amount of SARS-CoV-2 shed per g (S_w) was determined using

$$S_w = \frac{C_w F_w}{M_f N_i} \quad (1)$$

where C_w is the N2 gene concentration (copies per L), F_w is the wastewater flow (L per day), M_f is the mass of feces excreted by a person (g per day), and N_i is the number of SARS-CoV-2 infected individuals.[22]

Bayesian models. Mixed effect models relating observed N2 concentrations to observed sewershed cases normalized to a population of 100 000 individuals were generated using R-Statistics (version 4.4.1), rstan (2.36.6), brms (2.21.0), bayesplot (1.11.1), and cross-validation comparisons were performed using loo (2.8.0). The following models were generated:

Model 1 – random intercept to account for WWTP site clustering with a flow-adjusted N2 fixed effect predictor:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{0j} + \varepsilon_{ij} \quad (2)$$

Model 2 – random intercept to account for WWTP site clustering as well as random slopes of flow-adjusted N2 to account for heterogeneity at different WWTP locations.

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{0j} + b_{1j} x_{ij} + \varepsilon_{ij} \quad (3)$$

Model 3 – random intercept to account for WWTP site clustering as well as random slopes of flow-adjusted N2 with an interaction term to allow the slopes to vary by VOC categories (alpha, delta and omicron) using pre-VOC as a reference.

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_{ij} + \beta_3 (x_{ij} \cdot z_{ij}) + b_{0j} + b_1 x_{ij} + \varepsilon_{ij} \quad (4)$$

Where

y_{ij} is the i-th observation of the response variable (*i.e.*, log10 of the COVID-19 cases per 100 000 individuals) for the j-th group, where group is sewershed location

β_0 is the fixed intercept or the overall mean effect across all sewersheds

β_1 is the fixed slope for the predictor x_{ij} (*i.e.*, log_n2_flowadj), representing its effect on the response variable

β_2 is the fixed slope for the predictor z_{ij} (*i.e.*, the SARS-CoV-2 VOC), representing its effect on the response variable

β_3 is the fixed slope for the $x_{ij} \cdot z_{ij}$, representing the interactions of log_n2_flowadj and SARS-CoV-2 VOC on the response variable

b_{0j} is the random intercept for the j-th sewershed, representing specific differences from the overall intercept

b_{1j} is the random slope for the predictor x_{ij} (*i.e.*, log_n2_flowadj) in the j-th sewershed

ε_{ij} is the residual error (*i.e.*, unexplained variability) for the i-th observation of the response in the j-th sewershed.

Details of the models and results are in the ESI† 1.

Results & discussion

Wastewater N2 concentrations and reported cases

We tracked wastewater N2 gene fragment concentrations at eleven sampling sites in southwest Ohio (Fig. 1 and 2). Consistent with general trends across the United States, we found there to be a large increase in cases during the winter of 2020–2021, and subsequent waves of differing sizes associated with the introductions of the alpha, delta, and omicron variants (Fig. 2A and B). We generally found close correlations between the raw N2 concentrations (copies per L) and a 7 day rolling average (3 days before, 3 days after) of reported cases (Pearson's R ranging from 0.4 to 0.83, with one outlier that was likely due to a rain event, Fig. 2B). However, there were nuances in the relationship between N2 concentrations and case counts over different phases of the pandemic, which are further described below.

Many studies have explored whether adjusting SARS-CoV-2 viral gene copies to account for changes in wastewater strength (*i.e.*, due to varying levels of dilution) might help wastewater measurements better reflect actual trends in community infection conditions. Adjustment methods tested include quantification of endogenous markers like *Carjivirus* (CrAssphage) or pepper mild mottle virus to represent fecal loading in the sample[13,23] and using physical or chemical parameters like total suspended solids, conductivity, and biological oxygen demand.[24,25] Evaluating whether such adjustments improve the robustness of wastewater measurements as a public health indicator is generally based on whether these adjustments increase the correlation between wastewater signal and reported cases. Unfortunately, there is not a strong consensus in findings, but multiple studies have reported that the use of flow is as good as or better than other proposed metrics[9,13,25] at improving the correlation. Our study too found that using total daily flow to adjust the wastewater concentrations (resulting in a value reflecting N2 copies per day at the sampling site, as opposed to the original measurement N2 copies per L in the individual sample) increased the correlation between N2 and case counts at every single site (Pearson's R ranging from 0.62 to 0.87) (Fig. 2C and D).

Variants of concern in wastewater samples

We sequenced samples from January 2021 onward since this corresponded with the arrival of the alpha variant to the

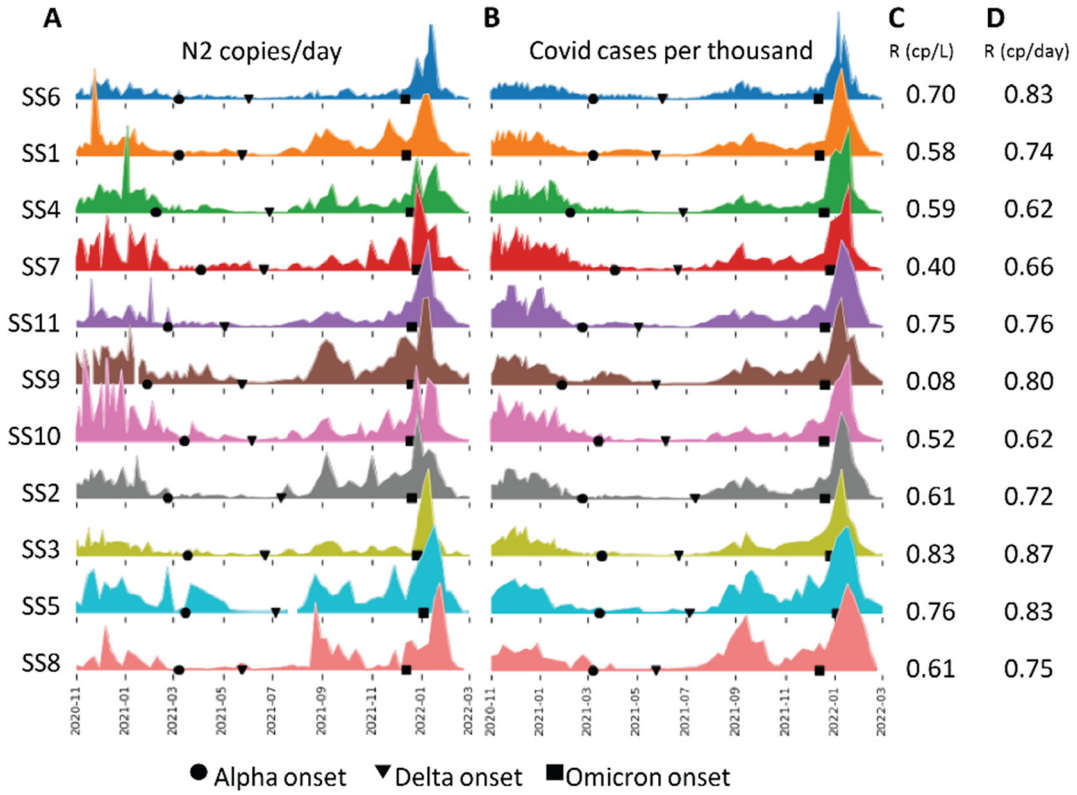

Fig. 2 (A) N2 gene fragment copies per day at each sampling site between November 2020 and March 2022. *Y*-axis scale differs for each sampling site and is not shown here due to size; larger versions with axis labels are available in Fig. S1† (N2 copies per L) and Fig. S2† (N2 copies per day) and (B) daily COVID-19 cases per thousand (*Y*-axis range is 0 to 3.62 for all sampling sites; larger version available in Fig. S1 and S2†) in each of the contributing sewersheds. Sites are ordered by size of contributing population, from largest (SS6) to smallest (SS8); refer to Fig. 1 for site codes. Symbols show first detection of alpha (circle), delta (triangle), and omicron (square) variants of concern in the wastewater sequencing data from this study. (C) Pearson's correlation between unadjusted N2 load (copies per L) and case counts. (D) Pearson's correlation between flow-adjusted N2 load (copies per day) and case counts.

United States. It was clear that the alpha, delta, and omicron variants appeared and dominated in succession at all locations. Some of the less abundant VOCs in the United States, like gamma, epsilon, iota, and mu, were also detected, though not at every sampling site (Fig. 3). While each of the three VOC waves we tracked roughly co-occurred at all sampling sites, the week of first appearance differed between sites.

In addition to sequencing, we examined the spring 2021 alpha variant wave in greater detail using ddPCR mutation assays for the N501Y and Δ H69/70 deletion, both of which were considered defining of the alpha variant (Fig. S1†). Others have successfully used targeted approaches to identify incoming VOCs, including a nationwide reporting system in Uruguay[4] and a retrospective analysis of municipal wastewater.[14] The use of ddPCR assays allows for rapid data analysis within a shorter timeframe as opposed to high-throughput sequencing, which is more laborious. Therefore, we assessed the performance of the ddPCR assays compared to the results from our sequenced libraries. In October and November 2020, 100% of the SARS-CoV-2 RNA sequences detected were non-alpha, and no sequences associated with the ΔH69/70 deletion were detected using the ddPCR assay. Alpha sequences were first detected in wastewater on January 28, 2021, in SS9 and were present in all sewersheds by March 18, 2021 (Table S1†). Between March and May 2021, the proportion of sequences associated with ΔH69/70 increased to nearly 90% at all sampling sites (Fig. S3†), indicating alpha was likely the dominant variant in the sampling areas by mid-late April. This timeline was also reflected in the sequencing data.

At our sampling locations, alpha took an average of 7.8 weeks to comprise greater than 95% of sequences, longer than delta (which took 4 weeks on average) and Omicron (average 3.6 weeks; Table S1†). A previous study conducted in Catalonia found that the alpha variant became dominant in wastewater in an average of 11 weeks from its first detection.[26] Clinical genomic surveillance demonstrated that the delta variant had a greater logistic growth rate and effective reproduction number than alpha,[27] leading to a more rapid rise to dominance, but with geographic variation in transmission advantage. This may explain why the variants rise to dominance at different rates. Trends in wastewater detection do not always correspond temporally with clinical detection; for example, a mutation profile associated with the delta variant was detected in wastewater in Bangladesh as

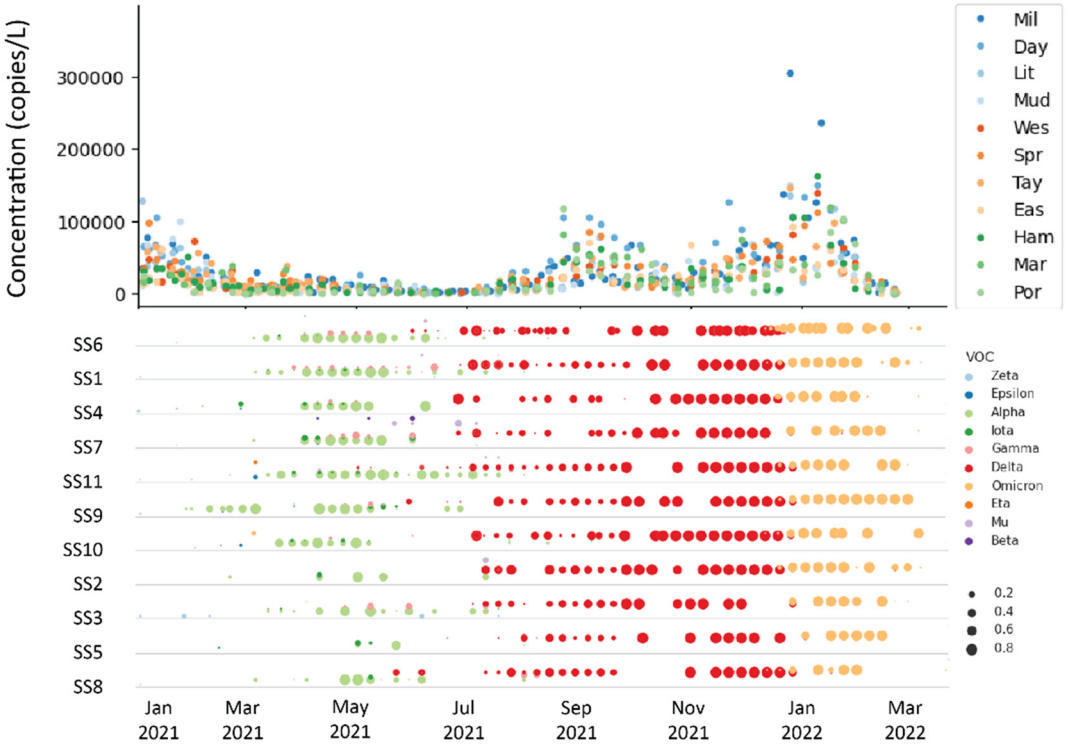

Fig. 3 N2 concentrations (copies per L) at each of the sampling sites, and proportion of total sequences in a sample assigned to each VOC (represented by size of dot), as assigned by Freyja. Gaps in timeline indicate either absence of VOC or lack of sequencing data for that time point. Sites are ordered by size of contributing population, from largest (SS6) to smallest (SS8); refer to Fig. 1 for site codes.

early as October 2020,[28] whereas its first clinical sequence was not reported until spring of 2021, when it was first identified as a VOC in India.

Other mutations

In addition to tracking the dynamics of VOCs based on their defining mutations (which are not always subject to a full consensus), wastewater sequencing data allows for the detection of all mutations and can reveal additional dynamics occurring in the population represented. For example, we found that the delta VOC completely dominated at all locations (>95%) between approximately July and December of 2021 (Fig. 3). This assessment, using the open-source tool Freyja, is based on WHO designations for the set of Spike gene mutations used to define each VOC.[17] However, we found 8760 additional mutations that reached a frequency of at least 10% at a minimum depth of 30×; 273 of these reached a frequency of at least 95%. Some of these mutations were detectable in our samples prior to the emergence of any VOC (*i.e.*, prior to January 2021), indicating that variants were circulating even in the fall, but none came to dominance until the arrival of alpha. Other mutations occurred at approximately the same times as various VOCs, despite not being considered defining mutations of those VOCs by the WHO. For example, genomes classified as belonging to the delta variant is defined by a specific set of mutations, including G24410A (Fig. 4A). The mutations C22227T, C24208T, and C25339T were detected in the summer and fall of 2021 (corresponding with the delta wave) but did not achieve and maintain 100% alternative allele frequency in the way that delta-defining mutations did (Fig. 4B). We examined their global incidence using data available from genomes uploaded to NextStrain and found that each of these mutations was only found in genomes in the delta clade (Fig. 4A and C–E). However, while these mutations did co-occur temporally at our sampling sites (Fig. 4B), they did not co-occur with one another in any genomes within NextStrain's clinical genome dataset (Fig. 4C–E).

There did appear to be some geographic specificity as well; for example, a mutation at genome site 25339 was observed in samples from the Cincinnati area during spring 2021 (March through June) but did not appear at in other samples until after July. This mutation has only been associated with the delta clade and reported globally at frequencies greater than 1% only after June 2021, so its appearance earlier in the spring at certain sampling sites is a demonstration of mutations spreading in the population that are largely undetected in clinical studies.

Fecal load

Sampling design continues to be an open topic of study for wastewater surveillance studies, particularly whether to sample at the WWTP level, capturing a large population, or focus on sub-sewersheds. In our study, the size of the

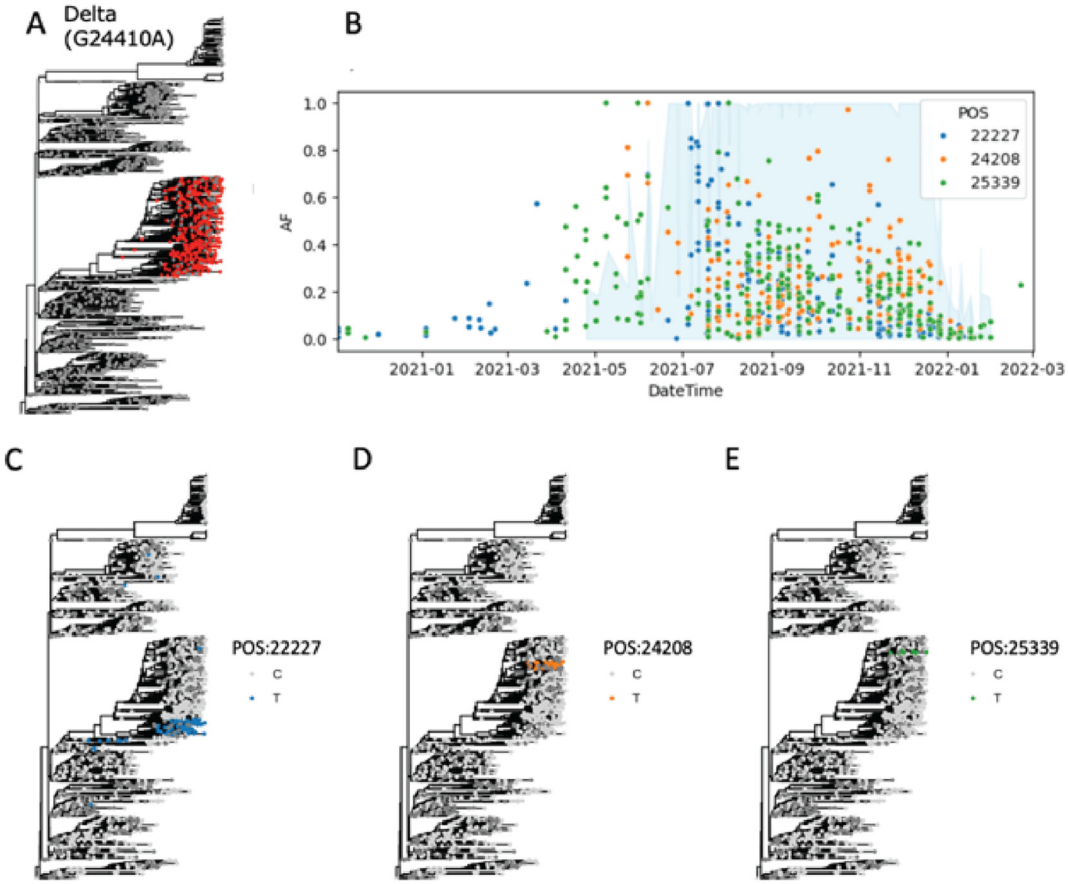

Fig. 4 (A) Tree of SARS-CoV-2 genomes from NextStrain with delta clade colored red. Only genomes uploaded before April 2022 are included. (B) Frequency of delta VOC (blue shading) as well as three "non-defining" mutations associated with subclades. Frequency of delta is represented by the allele frequency of G24410A, considered a defining mutation for the VOC. Frequencies for the mutations C22227T, C24208T, and C25339T across the time series are shown in blue, orange, and green respectively. (C–E) Presence of the above mutations in genomes on the NextStrain tree.

population within the sewershed was not associated with a stronger or weaker correlation between N2 load and reported case counts (Pearson's $R < 0.2$). Other studies have reported the SARS-CoV-2 viral shedding rate per gram of feces based on the known sewershed population and a previously estimated value of feces per person, 128 g.[29] Applying eqn (1), we found an average fecal load of approximately 8.67 $\log_{10}$ copies per g feces over all samples. This was similar to a study in Arizona conducted over a comparable time period, which found fecal loads falling within the range of 7.53–9.29 $\log_{10}$ gene copies per g feces.[30] However, much like this study, we found that there were sampling site-specific differences that would require further investigation to understand (Fig. S4 and S5†). Specifically, SS1's 95% credible interval was different from all the other sewersheds (ESI† 1). An underestimate of the population or an underreporting of case counts could be a possible explanation.

The calculation for estimating fecal load does not account for two important features of fecal shedding: (1) that not all infected individuals shed the virus in their feces, and (2) that individuals who do shed virus emit different amounts per gram of feces and do so for variable lengths of time. A clinical study where N2 was quantified from the dry weight feces of positive individuals found up to 6 $\log_{10}$ copies per mg, *i.e.* 9 $\log_{10}$ copies per g feces shed by individuals.[31] Shedding persisted for some up to 30 days after symptom onset, with a very wide range of shedding concentrations and temporal trajectories. On a larger scale, the uncertainty and variation in population transience, combined *vs.* separated sewers, rainfall, loss of signal, and many other factors could impact this value.[32] Prasek *et al.*[30] found community-specific differences in shedding rates and suggested demographics like age and ethnicity may influence these.

Different phases of the pandemic

The ability of reported cases to represent actual SARS-CoV-2 prevalence depends on the extent of testing availability and occurrence whereas wastewater surveillance is, in theory, agnostic to such variation. However, because different variants have emerged and dominated throughout the course of the pandemic, there is also the question of whether fecal shedding rates remain consistent. Clinical studies have characterized differences in nasopharyngeal shedding between the omicron and delta variants, with delta-infected patients generally having greater infectious

Table 2 Performance metrics of the models. See Methods for description of models. SE = standard error

Models	Expected log predictive density (SE)	Effective number of parameters (SE)	Leave-one-out information criterion (SE)	Expected log predictive density differences (SE)
1	−263.4 (26.4)	14.1 (1.1)	526.9 (52.7)	−139.6 (15.6)
2	−250.3 (26.5)	23.1 (2.5)	500.6 (53.0)	−126.4 (15.1)
3	−123.9 (26.4)	28.4 (3.0)	247.5 (52.9)	0.0 (0.0)

loads and differential patterns of shedding.[33] In a wastewater context, Prasek *et al.*[30] observed an increase in estimated shedding rate following the emergence of the delta variant. In our Bayesian analysis, alpha phase showed the highest fecal load, followed by delta, then pre-VOC, and finally omicron (see ESI† 1). There are at least two possible explanations for this observation: 1) it is an artifact of lower testing rates, which would affect the denominator in eqn (1) and (2) alpha variant presented a higher fecal shedding rate compared to other variants. Radu *et al.*[34] observed that the arrival of the alpha variant in Vienna, Austria, corresponded with an increase in wastewater SARS-CoV-2 concentrations despite a decline in clinical case counts, implying increased fecal shedding of the virus in alpha. Several other studies have also shown that wastewater samples from the time period associated with the omicron variant indicated lower fecal shedding rates.[25,30] When interpreting wastewater surveillance data to estimate SARS-CoV-2 variant shedding it is important to consider if reported cases represent actual SARS-CoV-2 prevalence.

We applied three Bayesian linear mixed-effects models to our time series data to evaluate how well the N2 load could predict reported COVID-19 cases across the WWTP sites while accounting for site-specific differences. Model 1 incorporates N2 load and WWTP site as fixed and random effects, respectively. Models 2 and 3 both incorporate N2 load and WWTP site as random effects, accounting for slope and site-specific differences. Model 3, however, allows N2 load to interact with the phase of the pandemic (*i.e.* which VOC was dominant) to reveal the influence of VOC on slope differences for each site. Leave-one-out information cross validation analysis indicated that model 3 outperformed model 1 and model 2 (Table 2, see ESI† 1) because it accounted for WWTP clustering and the heterogeneity of N2 slopes for the VOC categories (*i.e.*, preVOC, alpha, delta, omicon). Posterior predictive analysis comparing the observed COVID-19 cases with the model 3's 95% prediction credible intervals indicated good model fit (Fig. 5). Plots of the posterior distributions demonstrated that the intercepts for SS1 and SS4 were credibly lower than the grand mean

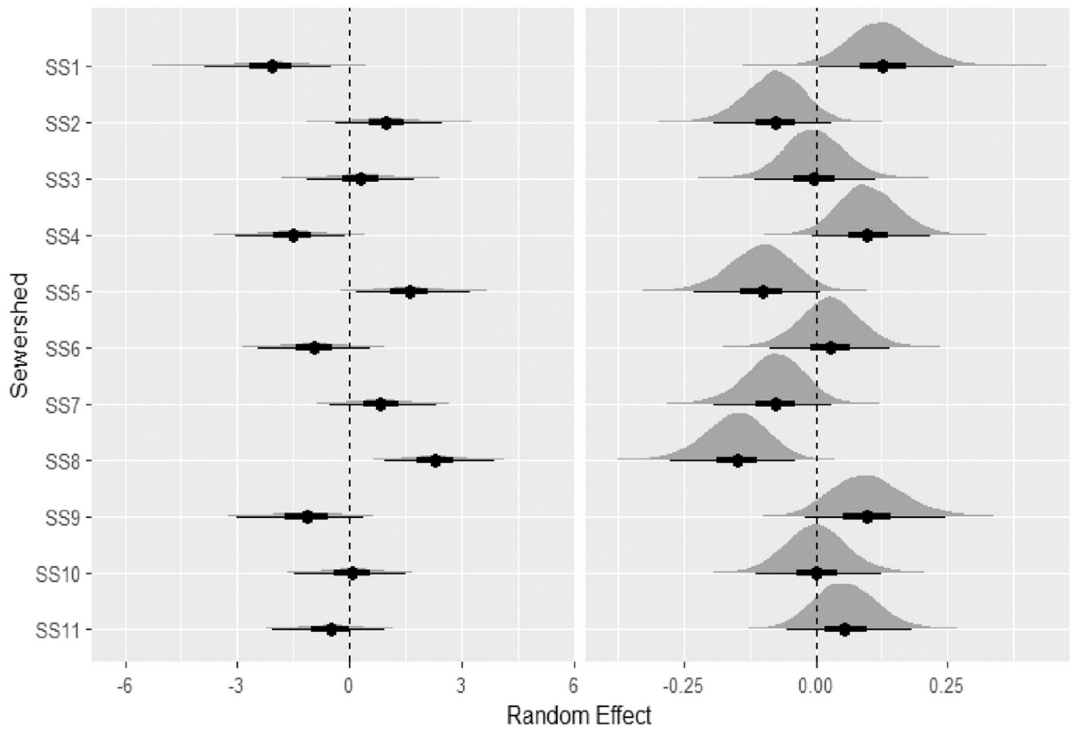

Fig. 5 Top performing Bayesian model 3, which accounts for sewershed and SARS-CoV-2 variant heterogeneity in the relationship between daily N2 loads in wastewater and COVID-19 cases per 100 000 individuals. (left side) Intercept MCMC intervals and (right side) slope MCMC intervals, where the inner and outer error bars represent the 50 and 95% credible intervals.

across all WWTP sites, while the slope of SS1 was credibly higher than the overall average slope (Fig. 5). By contrast, SS5 and SS8 had higher intercepts than the grand mean across all WWTP sites, while the slope of SS8 was credibly lower than the overall average slope. The intercepts and slopes of the other WWTPs, while not credibly different from the overall averages using 95% credibility threshold, still exhibited a similar pattern. The reciprocal relationship between the intercepts and slopes suggests that populations with more severe outbreaks tend to stabilize or decrease more rapidly, whereas those with less severe outbreaks are more likely to increase more quickly over time. This could be due to a number of factors, such as differences in virus fitness and transmissibility, in fecal shedding, in host immunity, distinct demographics contributing to the sewersheds, differential case reporting at different sampling sites, and more.

We found that the model 3 better predicted case counts than the other models where phase was not included as a factor. By including random effects for both the phase-specific slopes and the WWTP locations, model 3 captured how baseline infection levels differ between locations and how the effect of different SARS-CoV-2 variants on COVID-19 infections varies depending on the location. Model 3 demonstrated credible variation across different WWTP locations, both in baseline COVID-19 cases and strength of the association between phase-specific effects and COVID-19 cases.

Conclusion

We analyzed quantitative N2 concentrations and qualitative sequencing data for SARS-CoV-2 in wastewater from 11 sampling sites in Southwestern Ohio. Over the course of our time series, we were able to detect the emergence of the three main VOCs, alpha, delta, and omicron, in addition to the presence of other less abundant VOCs. We found that flow-adjusted N2 concentrations correlated strongly with reported COVID-19 cases, but that the relationship between N2 load and number of cases varied during distinctive phases of the pandemic. This finding supports the concept that the different variants of SARS-CoV-2 have diverse levels and lengths of fecal shedding. Multi-level modeling underscores the importance of accounting for sampling site-specific heterogeneity to understand the trajectory of COVID-19 infections monitored in sewersheds.

Data availability

The data supporting this article have been included as part of the ESI.†

Disclaimer

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. The views expressed in this paper are those of the authors and do not necessarily represent the views or the policies of the U.S. Environmental Protection Agency. Mention of trade names, products, or services does not convey, and should not be interpreted as conveying, official EPA approval, endorsement, or recommendation.

GitHub repository for code

https://github.com/USEPA/Dynamics-of-SARS-CoV-2

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank the Ohio Department of Health and the Ohio Wastewater Monitoring Network for coordination of sample collection, data analysis and public reporting. We thank the wastewater utilities for providing samples. We thank Pegasus Technical Services (specifically Barry Wiechman, Anna Braam, Sara Okum and Raghu Venkatapathy) for technical support. This research was funded by EPA Office of Research and Development's internal research program.

References

1 C. L. Swift, M. Isanovic, K. E. Correa Velez, S. C. Sellers and R. S. Norman, Wastewater surveillance of SARS-CoV-2 mutational profiles at a university and its surrounding community reveals a 20G outbreak on campus, *PLoS One*, 2022, **17**(4), e0266407.

2 V. Vo, R. L. Tillett, C. L. Chang, D. Gerrity, W. Q. Betancourt and E. C. Oh, SARS-CoV-2 variant detection at a university dormitory using wastewater genomic tools, *Sci. Total Environ.*, 2022, **805**, 149930.

3 Y. Wang, P. Liu, H. Zhang, M. Ibaraki, J. VanTassell and K. Geith, *et al.* Early warning of a COVID-19 surge on a university campus based on wastewater surveillance for SARS-CoV-2 at residence halls, *Sci. Total Environ.*, 2022, **821**, 153291.

4 F. Cancela, N. Ramos, D. S. Smyth, C. Etchebehere, M. Berois and J. Rodriguez, *et al.* Wastewater surveillance of SARS-CoV-2 genomic populations on a country-wide scale through targeted sequencing, *PLoS One*, 2023, **18**(4), e0284483.

5 A. Crits-Christoph, R. S. Kantor, M. R. Olm, O. N. Whitney, B. Al-Shayeb and Y. C. Lou, *et al.*, Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants, *MBio*, 2021, **12**(1), DOI: 10.1128/mbio.02703-20.

6 H. Wani, S. Menon, D. Desai, N. D'Souza, Z. Bhathena and N. Desai, *et al.* Wastewater-Based Epidemiology of SARS-CoV-2: Assessing Prevalence and Correlation with Clinical Cases, *Food Environ. Virol.*, 2023, **15**(2), 131–143.

7 S. G. Rimoldi, F. Stefani, A. Gigantiello, S. Polesello, F. Comandatore and D. Mileto, *et al.* Presence and infectivity of SARS-CoV-2 virus in wastewaters and rivers, *Sci. Total Environ.*, 2020, **744**, 140911.

8 A. Y. Rocha, M. E. Verbyla, K. E. Sant and N. Mladenov, Detection, Quantification, and Simplified Wastewater Surveillance Model of SARS-CoV-2 RNA in the Tijuana River, *ACS ES&T Water*, 2022, **2**(11), 2134–2143.
9 Y. Ai, A. Davis, D. Jones, S. Lemeshow, H. Tu and F. He, *et al.* Wastewater SARS-CoV-2 monitoring as a community-level COVID-19 trend tracker and variants in Ohio, United States, *Sci. Total Environ.*, 2021, **801**, 149757.
10 L. Hopkins, D. Persse, K. Caton, K. Ensor, R. Schneider and C. McCall, *et al.* Citywide wastewater SARS-CoV-2 levels strongly correlated with multiple disease surveillance indicators and outcomes over three COVID-19 waves, *Sci. Total Environ.*, 2023, **855**, 158967.
11 A. Nemudryi, A. Nemudraia, T. Wiegand, K. Surya, M. Buyukyoruk and C. Cicha, *et al.* Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater, *Cell Rep. Med.*, 2020, **1**(6), 100098.
12 Z. Bohrerova, N. E. Brinkman, R. Chakravarti, S. Chattopadhyay, S. A. Faith and J. Garland, *et al.* Ohio Coronavirus Wastewater Monitoring Network: Implementation of Statewide Monitoring for Protecting Public Health, *J. Public Health Manag. Pract.*, 2023, **29**(6), 845–853.
13 M. Nagarkar, S. P. Keely, M. Jahne, E. Wheaton, C. Hart and B. Smith, *et al.* SARS-CoV-2 monitoring at three sewersheds of different scales and complexity demonstrates distinctive relationships between wastewater measurements and COVID-19 case data, *Sci. Total Environ.*, 2022, **816**, 151534.
14 S. Agrawal, L. Orschler, S. Schubert, K. Zachmann, L. Heijnen and S. Tavazzi, *et al.* Prevalence and circulation patterns of SARS-CoV-2 variants in European sewage mirror clinical data of 54 European cities, *Water Res.*, 2022, **214**, 118162.
15 M. F. Smith, S. C. Holland, M. B. Lee, J. C. Hu, N. C. Pham and R. A. Sullins, *et al.* Baseline Sequencing Surveillance of Public Clinical Testing, Hospitals, and Community Wastewater Reveals Rapid Emergence of SARS-CoV-2 Omicron Variant of Concern in Arizona, USA, *MBio*, 2023, **14**(1), e0310122.
16 S. Agrawal, L. Orschler, S. Tavazzi, R. Greither, B. M. Gawlik and S. Lackner, Genome Sequencing of Wastewater Confirms the Arrival of the SARS-CoV-2 Omicron Variant at Frankfurt Airport but Limited Spread in the City of Frankfurt, Germany, in November 2021, *Microbiol. Resour. Announce.*, 2022, **11**(2), e0122921.
17 S. Karthikeyan, J. I. Levy, P. De Hoff, G. Humphrey, A. Birmingham and K. Jepsen, *et al.* Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission, *Nature*, 2022, **609**(7925), 101–108.
18 N. Vega-Magana, R. Sanchez-Sanchez, J. Hernandez-Bello, A. A. Venancio-Landeros, M. Pena-Rodriguez and R. A. Vega-Zepeda, *et al.* RT-qPCR Assays for Rapid Detection of the N501Y, 69-70del, K417N, and E484K SARS-CoV-2 Mutations: A Screening Strategy to Identify Variants With Clinical Impact, *Front. Cell. Infect. Microbiol.*, 2021, **11**, 672562.
19 Illumina, NextSeq System Denature and Dilute Libraries Guide Document #15048776 v09, 2018.
20 N. D. Grubaugh, K. Gangavarapu, J. Quick, N. L. Matteson, J. G. De Jesus and B. J. Main, *et al.* An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar, *Genome Biol.*, 2019, **20**(1), 8.
21 A. Wilm, P. P. Aw, D. Bertrand, G. H. Yeo, S. H. Ong and C. H. Wong, *et al.* LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets, *Nucleic Acids Res.*, 2012, **40**(22), 11189–11201.
22 K. Curtis, D. Keeling, K. Yetka, A. Larson and R. Gonzalez, Wastewater SARS-CoV-2 RNA concentration and loading variability from grab and 24-hour composite samples, *medRxiv*, 2021, preprint, DOI: **10.1101/2020.07.10.20150607**.
23 H. D. Greenwald, L. C. Kennedy, A. Hinkle, O. N. Whitney, V. B. Fan and A. Crits-Christoph, *et al.* Tools for interpretation of wastewater SARS-CoV-2 temporal and spatial trends demonstrated with data collected in the San Francisco Bay Area, *Water Res.: X*, 2021, **12**, 100111.
24 J. M. Hutchison, Z. Li, C. N. Chang, Y. Hiripitiyage, M. Wittman and B. S. M. Sturm, Improving correlation of wastewater SARS-CoV-2 gene copy numbers with COVID-19 public health cases using readily available biomarkers, *FEMS Microbes*, 2022, **3**, xtac010.
25 R. Schill, K. L. Nelson, S. Harris-Lovett and R. S. Kantor, The dynamic relationship between COVID-19 cases and SARS-CoV-2 wastewater concentrations across time and space: Considerations for model training data sets, *Sci. Total Environ.*, 2023, **871**, 162069.
26 A. Carcereny, D. Garcia-Pedemonte, A. Martinez-Velazquez, J. Quer, D. Garcia-Cehic and J. Gregori, *et al.*, Dynamics of SARS-CoV-2 Alpha (B.1.1.7) variant spread: The wastewater surveillance approach, *Environ. Res.*, 2022, **208**, 112720.
27 R. Earnest, R. Uddin, N. Matluk, N. Renzette, S. E. Turbett and K. J. Siddle, *et al.* Comparative transmissibility of SARS-CoV-2 variants Delta and Alpha in New England, USA, *Cell Rep. Med.*, 2022, **3**(4), 100583.
28 R. Haque, M. E. Hossain, M. Miah, M. Rahman, N. Amin and Z. Rahman, *et al.* Monitoring SARS-CoV-2 variants in wastewater of Dhaka City, Bangladesh: approach to complement public health surveillance systems, *Hum. Genomics*, 2023, **17**(1), 58.
29 C. Rose, A. Parker, B. Jefferson and E. Cartmell, The Characterization of Feces and Urine: A Review of the Literature to Inform Advanced Treatment Technology, *Crit. Rev. Environ. Sci. Technol.*, 2015, **45**(17), 1827–1879.
30 S. M. Prasek, I. L. Pepper, G. K. Innes, S. Slinski, W. Q. Betancourt and A. R. Foster, *et al.* Variant-specific SARS-CoV-2 shedding rates in wastewater, *Sci. Total Environ.*, 2023, **857**(Pt 1), 159165.
31 P. J. Arts, J. D. Kelly, C. M. Midgley, K. Anglin, S. Lu and G. R. Abedi, *et al.* Longitudinal and quantitative fecal shedding dynamics of SARS-CoV-2, pepper mild mottle virus, and crAssphage, *mSphere*, 2023, **8**(4), e0013223.

32 C. Wiesner-Friedman, N. E. Brinkman, E. Wheaton, M. Nagarkar, C. Hart and S. P. Keely, *et al.* Characterizing Spatial Information Loss for Wastewater Surveillance Using crAssphage: Effect of Decay, Temperature, and Population Mobility, *Environ. Sci. Technol.*, 2023, **57**(49), 20802–20812.
33 O. Puhach, B. Meyer and I. Eckerle, SARS-CoV-2 viral load and shedding kinetics, *Nat. Rev. Microbiol.*, 2023, **21**(3), 147–161.
34 E. Radu, A. Masseron, F. Amman, A. Schedl, B. Agerer and L. Endler, *et al.* Emergence of SARS-CoV-2 Alpha lineage and its correlation with quantitative wastewater-based epidemiology data, *Water Res.*, 2022, **215**, 118257.