Science of the Total Environment 877 (2023) 162862

Contents lists available at ScienceDirect

Science of the Total Environment

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

Wastewater surveillance uncovers regional diversity and dynamics of SARS-CoV-2 variants across nine states in the USA

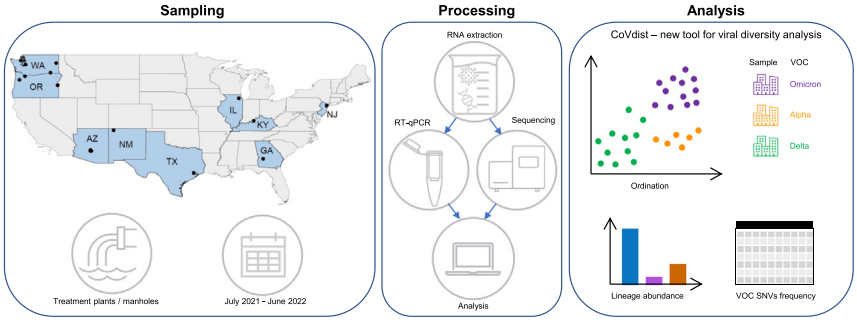
Rafaela S. Fontenele [a], Yiyan Yang [a], Erin M. Driver [b], Arjun Magge [b], Simona Kraberger [c], Joy M. Custer [c], Keith Dufault-Thompson [a], Erin Cox [b], Melanie Engstrom Newell [b], Arvind Varsani [c,d,e], Rolf U. Halden [b,f,g], Matthew Scotch [b,h], Xiaofang Jiang [a,*]

[a] National Library of Medicine, National Institute of Health, 8600 Rockville Pike, Bethesda, MD 20894, USA
[b] Biodesign Center for Environmental Health Engineering, Biodesign Institute, Arizona State University, Tempe, AZ 85281, USA
[c] The Biodesign Center for Fundamental and Applied Microbiomics, Arizona State University, Tempe, AZ 85287, USA
[d] School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA
[e] Center of Evolution and Medicine, Arizona State University, Tempe, AZ 85287, USA
[f] School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85281, USA
[g] OneWaterOneHealth, Nonprofit Project of the Arizona State University Foundation, Tempe, AZ 85287, USA
[h] College of Health Solutions, Arizona State University, Phoenix, AZ 85004, USA

HIGHLIGHTS

- Clear shifts between the Delta and Omicron lineages were seen through WBE samples.
- CoVdist is a new distance metric tool for calculating viral diversity in wastewater.
- CoVdist enables wastewater analysis in the context of the VOCs genetic diversity.
- Lineage abundance varied significantly between neighborhood and city scale samples.
- Recombinant lineages were observed during the transition from Delta to Omicron.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Damià Barceló

Keywords:
Wastewater-based epidemiology (WBE)
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
Coronavirus infectious disease 19 (COVID-19)

ABSTRACT

Wastewater-based epidemiology (WBE) is a non-invasive and cost-effective approach for monitoring the spread of a pathogen within a community. WBE has been adopted as one of the methods to monitor the spread and population dynamics of the SARS-CoV-2 virus, but significant challenges remain in the bioinformatic analysis of WBE-derived data. Here, we have developed a new distance metric, CoVdist, and an associated analysis tool that facilitates the application of ordination analysis to WBE data and the identification of viral population changes based on nucleotide variants. We applied these new approaches to a large-scale dataset from 18 cities in nine states of the USA using wastewater collected from July 2021 to June 2022. We found that the trends in the shift between the Delta and Omicron SARS-CoV-2 lineages were largely consistent with what was seen in clinical data, but that wastewater analysis offered the added benefit of revealing significant differences in viral population dynamics at the state, city, and even neighborhood scales. We also were able to observe the early spread of variants of concern and the presence of recombinant lineages during the transitions between variants, both of which are challenging to analyze based on clinically-derived viral genomes. The methods outlined here will be beneficial for future applications of WBE to monitor SARS-CoV-2, particularly as clinical monitoring becomes less prevalent. Additionally, these approaches are generalizable, allowing them to be applied for the monitoring and analysis of future viral outbreaks.

* Corresponding author.
E-mail address: xiaofang.jiang@nih.gov (X. Jiang).

http://dx.doi.org/10.1016/j.scitotenv.2023.162862
Received 7 January 2023; Received in revised form 9 March 2023; Accepted 10 March 2023
Available online 16 March 2023

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in late 2019 and resulted in the global COVID-19 pandemic. To inform COVID-19 responses, surveillance tools such as clinical genomic epidemiology were promptly applied to track pathogen transmission, estimate infection cases, identify circulating variants, and study viral evolution. Subsequently, wastewater-based epidemiology (WBE), where public health indicators are monitored through the analysis of wastewater samples, was adopted to complement clinical-based genomic epidemiology on SARS-CoV-2 (Jones et al., 2020; Mantilla-Calderon et al., 2022; Park et al., 2021; Xiao et al., 2020; Xing et al., 2020). The key advantages of WBE are that it is low-cost, non-invasive, and provides an anonymous sampling opportunity that captures viral diversity from symptomatic and asymptomatic individuals within a community (Lavania et al., 2022; Park et al., 2021). The wide application of WBE for SARS-CoV-2 has played an important role in continuous monitoring at a community level, especially where clinical testing and sequencing efforts are insufficient.

Targeted high-throughput sequencing of wastewater samples makes it possible to obtain SARS-CoV-2 genomic data, allowing for the identification of single nucleotide variants (SNVs) in the viral population and thereby providing a way to monitor variants of concern in the sampled locations (Crits-Christoph et al., 2021; Fontenele et al., 2021; Gregory et al., 2021; Izquierdo-Lara et al., 2021; Jahn et al., 2022; Karthikeyan et al., 2022). Moreover, studies have shown that through WBE approaches, mutations can be identified before they are seen in clinically-derived sequencing data (Pérez-Cataluña et al., 2022), and variants of concern (VOC) can be detected before clinical cases (Jahn et al., 2022; Karthikeyan et al., 2022). So far, most of the studies in the USA have focused on just one or a few cities within a state (Karthikeyan et al., 2022; Rouchka et al., 2021) and have had relatively short collection timeframes limiting their utility in understanding long term trends in viral populations (Baaijens et al., 2022; Layton et al., 2022; Sutton et al., 2022; Swift et al., 2022; Vo et al., 2022).

Bioinformatic analysis of high-throughput sequencing data from wastewater is challenging because the sequencing data quality is often hindered by low genomic concentrations, primer bias during amplification, and fragmentation of the viral genome. In addition, the metagenomic nature of the wastewater sample, of which genetic material is derived from viruses of divergent evolution history within a population, creates challenges for accurate phasing and detection of specific circulating lineages. Since WBE is a developing field, various bioinformatic tools to address these issues are still being developed using different approaches, such as the co-occurrence of mutations associated with VOC lineages within reads (COJAC) (Jahn et al., 2022), sample deconvolution methods such as Freyja (Amman et al., 2022; Karthikeyan et al., 2022), and the comparison of mutations associated with specific lineages (Ellmen et al., 2021; Pechlivanis et al., 2022). However, the use of some standard metagenomic approaches, such as ordination analysis, are still underdeveloped in WBE research. Ordination analysis can summarize and present high-dimensional data sets in a low-dimensional ordination space, while capturing the similarity and dissimilarity of the original data. It allows us to visualize complex data and examine intra-group variability, and it could be a valuable tool for investigating trends in the COVID-19 pandemic.

In this study, we performed whole-genome amplicon sequencing of SARS-CoV-2 nucleic acids from 39 wastewater catchments across nine states and 18 cities in the USA, with the samples taken approximately twice per month from July 2021 to June 2022. We developed a distance measurement metric named CoVdist to evaluate the dissimilarity between the amplicon metagenomic samples and a bioinformatic framework that is suitable for SNV and indel calling from wastewater samples. We found that although the shift of Delta to Omicron SARS-CoV-2 lineage was consistent with the global trend during the time frame, there were significant regional variations, including the diversity of circulating VOC lineages, the time point that the frequency of the signature mutations for each VOC lineage changed, as well as the temporal dynamics of the relative abundance of the viral lineages. This study shows that WBE can potentially inform public health guidance and interventions on a city or even neighborhood scale.

2. Methods

2.1. Sample collection and processing

Samples were collected from 39 wastewater treatment plants or within-sewer collection system locations across nine states in the USA including Arizona, Georgia, Illinois, Kentucky, New Jersey, New Mexico, Oregon, Texas, and Washington. Arizona (number of catchments = 17) and Kentucky (number of catchments = 6) included neighborhood-scale manhole sampling locations (Supplementary Table 1). The sampling locations were variable with wastewater flows ranging from approximately 0.1–200 + million l/day and populations served of <1000 to 800,000 people. Composite wastewater samples were collected by automated high-frequency wastewater samplers deployed at each target location that were programmed to collect aliquots of wastewater over a 24-h period based on predefined flow or time requirements as determined by the participating municipality. Samples were collected at the plant headworks for wastewater treatment plants, while sewer collection system samples were taken at pump stations, permanent underground vault systems, or from target manhole locations. Collected samples were mixed well and transferred from the automated samplers to high density polyethene bottles for overnight shipment to Arizona State University (ASU) with a combination of blue and wet ice. Local samples were hand-delivered to ASU on wet ice on the same day samples were collected. Collection occurred approximately once every two weeks.

Samples were processed immediately upon receipt to limit RNA degradative losses. Approximately 70 ml of wastewater was vacuum filtered using a 0.45 μM polyethersulfone filter unit (Thermo Fisher Scientific, Waltham, MA) to remove larger debris. The filtrate was subsequently concentrated using Millipore Sigma Amicon Ultra Centrifugal Filter Units (Burlington, MA) with a 10,000 molecular weight cutoff filter and 15 ml holding capacity. Five sequential 20-min centrifugations were completed at 2200*g*. Total filtrate volumes passed through the centrifugal tubes and resultant concentrate volumes were recorded for calculation purposes. In total, 200 μl of the final concentrate were used for total RNA extraction using a Qiagen RNeasy Mini Kit (Hilden, Germany) following the manufacturer's specifications to a final volume of 50 μl.

2.2. SARS-CoV-2 RT-qPCR

Quantification of SARS-CoV-2 was performed using two assays, a singleplex (E gene target) and multiplex (N1, ORF1ab, and S gene targets). The singleplex targeting the E gene was designed and validated by Corman et al. (Corman et al., 2020). The probes for this assay were purchased from Integrated DNA Technologies (Coralville, IA), and the reaction was performed using Invitrogen SuperScript III Platinum One-Step qRT-PCR Kit (Carlsbad, CA). Thermal conditions were as follows: hold for 5 min at 50 °C, 2 min at 95 °C, followed by 40 cycles of 95 °C for 3 s and 58 °C for 30 s. The E gene standard curve was executed in triplicate (1×10^2–1×10^6 copies/μl) with a reaction efficiency of approximately 92 % and cycle threshold (Ct) ≤ 32.9. In the multiplex assay, the Applied BiosystemsTM TaqPathTM COVID-19 Combo Kit and TaqPathTM Multiplex Master Mix (No ROX) were used (Thermo Fisher Scientific - Waltham, MA). Manufacturer's recommendations were modified as follows (per well): 10 μl nuclease free water, 6.25 μl multiplex master mix, 2.5 μl MS2 phage control (diluted 1:10 with nuclease free water), and 1.25 μl COVID-19 real-time PCR assay multiplex, 5 μl of sample. Thermal conditions were as follows: hold for 2 min at 25 °C, 10 min at 53 °C, 2 min at 95 °C, followed by 40 cycles of 95 °C for 3 s and 60 °C for 30 s. The standard curve (1 × 101–1 × 106 copies/μl) was run in triplicate with reaction efficiencies of 100 % for N1 and S, and 96 % for ORF1ab; cycle threshold cut-off thresholds were ≤ 34.9, ≤ 31.9, and ≤ 34.9, respectively. The positive control was purchased from Twist Bioscience (Control 1, Australia/VIC01/

2020, MT007544.1). Negative controls were included in each sample batch, where deionized water went through the entire filtration, concentration, and RNA extraction steps. Additionally, no-template controls (nuclease free water) were included in each plate (1 well per 7 samples) to assess for contamination. All negative controls resulted in non-detects. Applied Biosystems QuantStudio 3 or 5 Real-Time PCR System (Foster City, CA, USA) were used for analysis, with Design and Analysis software version 1.5.1.

2.3. High-throughput sequencing

The extracted total RNA (16 μl) from each sample was used to generate cDNA using the Superscript® IV VILO Master Mix (ThermoFisher, Waltham, MA, USA) following the manufacturer's protocol with the reverse transcription incubation step (50 °C) for 30 min. 10 μl of cDNA from each sample was used to generate Illumina sequencing libraries with the xGen™ SARS-CoV-2 Amp Panel 96 rxn kits (IDT, Coralville, IA, USA). In addition, two controls were run per sequencing run (water control and a wastewater sample from 2019 before the SARS-CoV-2 pandemic to assess cross-contamination). The libraries were pooled (96 libraries that had both negative controls), normalized, and sequenced on an Illumina HiSeq 2500 sequencer (2 × 150 paired-end; 96 libraries per sequencing run) at Psomagen Inc. (Rockville, MD, USA).

2.4. Metagenomic data processing and analysis

We designed a SARS-CoV-2 metagenomics analysis pipeline to process the demultiplexed raw sequencing reads generated from the wastewater samples (Fig. 2A). The raw reads were received from Psomagen Inc. (Rockville, MD, USA) with the adapters removed. Briefly, the adapter-trimmed raw reads were aligned to the reference genome of SARS-CoV-2 (Wuhan-Hu-1/2019; MN908947; RefSeq ID NC_045512.2) using BWA-MEM (v. 0.7.17) (Li and Durbin, 2009). The amplification primers used for enrichment before sequencing were soft-clipped from the alignment using the tool "iVAR trim" (v.1.3.1) (Grubaugh et al., 2019), and the bam file was then re-aligned using "LoFreq viterbi" (v.2.1.5) (Wilm et al., 2012) to correct possible mapping errors. Subsequently, an initial variant call was performed using LoFreq (v.2.1.5) (Wilm et al., 2012) where the VCF file output was used to compute primer bias using iVAR. The flagged primer pairs are then used to remove primer-biased reads from the alignment files using iVAR tools (v.1.3.1) (Grubaugh et al., 2019). The final bam file with biased reads removed was then used for a final variant call using LoFreq with a minimum coverage of 5 reads, minimum quality base of 30 and minimum mapping score of 20. The VCF file was then annotated using snpEff (v5.0) (Cingolani et al., 2012) to obtain amino acid mutations. Only samples with >50 % of the genome covered with at least 5 reads per position were used for downstream analysis. The SARS-CoV-2 metagenomic data analysis pipeline is available in GitHub repository (https://github.com/nlm-irp-jianglab/bioinfo-wwbe).

2.5. Sample dissimilarity distance analysis with CoVdist

To calculate dissimilarity between SARS-CoV-2 in wastewater metagenomic samples, we proposed a new metric, CoVdist, based on the Yue & Clayton dissimilarity index (Yue and Clayton, 2005) to measure the pairwise distances between wastewater samples and lineages. Using the SNVs called from samples *i* and *j* mapped to a reference genome with a length of *N* bases, the CoVdist can be defined as:

$$CoVdist_{i,j} = \sum_{n=1}^{N}\left(1 - \frac{p_n^i \cdot p_n^j}{p_n^i \cdot p_n^i + p_n^j \cdot p_n^j - p_n^i \cdot p_n^j}\right)$$

where p_n^i is the vector representing the proportions of A, C, U/T, and G at position *n* for sample *i*. For example, a vector of [0.3, 0.4, 0.1, 0.2] indicates the allele frequencies of A, C, U/T and G are 30 %, 40 %, 10 % and 20 %, respectively. CoVdist is subject to the rules of distance: (1) The CoVdist between a sample and itself is always zero. (2) Its value is always non-negative; (3) It is symmetric; (4) It satisfies the triangle inequality.

CoVdist uses as an input the variant call file (VCF) and depth file from each sample. In this tool, the CoVdist matrix could be used to perform an ordination analysis through principal coordinate analysis (PCoA), multidimensional scaling (MDS) or t-distributed stochastic neighbor embedding (t-SNE) methods. In addition, we provide VCFs generated for each SARS-CoV-2 lineage based on the output of the phylogenetic tree UshER (Lanfear and Mansfield, 2020; Turakhia et al., 2021). The user can plot the lineages of interest along with the wastewater samples to identify the population diversity in the context of the genetic diversity of each SARS-CoV-2 lineage. The tool also contains an option to plot the results, generating an interactive plot in html format (https://github.com/nlm-irp-jianglab/CoVdist).

2.6. Identification of indels and SNVs associated with pangolin lineages

The SARS-CoV-2 genomes available at GISAID (Elbe and Buckland-Merrett, 2017; Shu and McCauley, 2017) on August 15th 2022 were downloaded for analysis. The genomes were then processed by Nextclade CLI (v.2.4.0) (Aksamentov et al., 2021) which generates a multiple sequence alignment against the reference genome GenBank accession # MN908947; RefSeq ID NC_045512.2 (Wuhan-Hu-1/2019) and provides a list of SNVs, insertions, and deletions associated with each genome sequence. The same set of genomes was analyzed with Phylogenetic Assignment of Named Global Outbreak Lineages (pangolin) (v.4.1.2-pdata-1.13) (O'Toole et al., 2021) to obtain a pangolin lineage classification for each genome. Only genomes that passed all quality controls as "good" applied by Nextclade and that passed pangolin quality control for lineage assignment were used for downstream analysis.

To identify SNVs associated with pangolin lineages and clades, a mutation annotated phylogenetic tree including all available genomes (n = 6080,78) from GISAID, GenBank, COG-UK, and CNCB generated by sarscov2phylo pipeline (v. 13-11-2020) used by the Ultrafast Sample placement on Existing tRee (UShER) (v. 0.5.6) (Lanfear and Mansfield, 2020; Turakhia et al., 2021) was downloaded on August 15th, 2022. The tool matUtils (McBroome et al., 2021) was used to extract (1) which mutations are associated with each node in the tree (2) the mutations associated with each lineage (from root to lineage) from the mutation annotated phylogenetic tree. The root to lineage mutations were then used to obtain a list of defining mutations per lineage, which consisted of those mutations associated with the lineage but not the ones present with their parental lineage.

2.7. VOC lineages relative abundance estimation in wastewater samples and clinical data

The lineage relative abundances for the wastewater samples were calculated by the tool Freyja (Karthikeyan et al., 2022). Clinical data of VOC lineage prevalence was obtained through the outbreak.info API which is enabled by GISAID (Gangavarapu et al., 2023). We focused on lineages belonging to the VOC lineages Alpha, Delta, and Omicron or recombinant lineages between Delta/Omicron or Omicron/Omicron lineages only. Prevalence was obtained for all nine states of the study in the same collection time frame from July 2021 to June 2022.

3. Results

3.1. Overview of the sequencing data from wastewater samples

In this study, we sequenced SARS-CoV-2 in 807 wastewater samples derived from 39 catchments in nine states in the USA: Arizona (sample size = 352), Georgia (sample size = 27), Illinois (sample size = 25), Kentucky (sample size = 77), New Jersey (sample size = 27), New Mexico (sample size = 26), Oregon (size = 77), Texas (sample size = 55) and Washington (sample size = 141) (Fig. 1A). The samples were collected

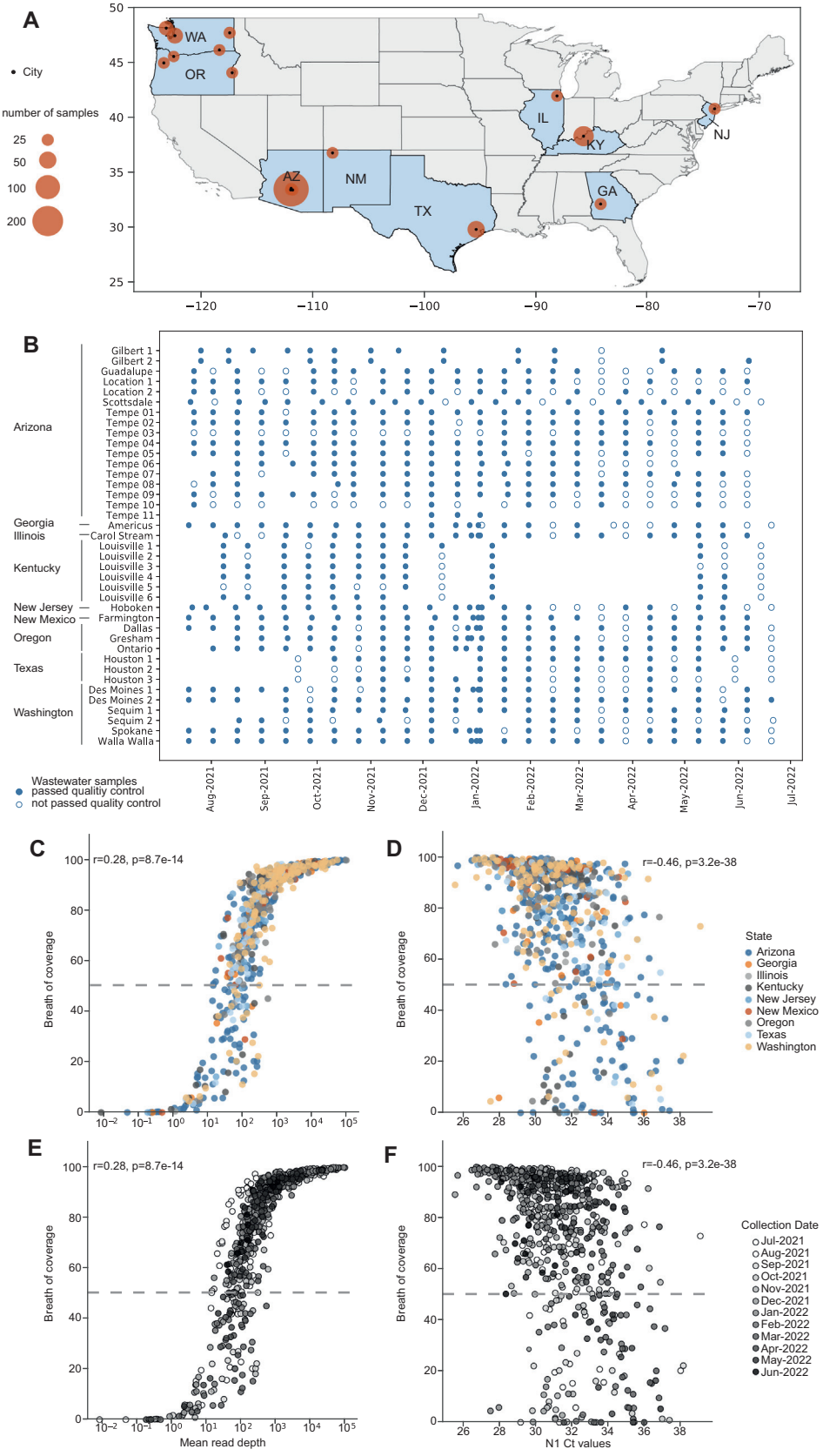


Fig. 1. Information on the wastewater samples from this study with sequencing metrics. A. Map showing the nine states where samples were collected, geographical coordinates of the cities where samples were collected, and the number of samples collected per city B. Graphical representation of the wastewater samples from the project by state/location and collection date. Filled circles represent samples that passed quality control, and empty circles samples represent those that did not pass quality control and were not used for the subsequent analyses. C. Correlation between breadth of coverage and mean depth of coverage and D. Correlation between breadth of coverage and N1 cycle threshold values (Ct values) with samples colored by state. E. Correlation between breadth of coverage and mean depth of coverage and F. correlation between breadth of coverage and N1 cycle threshold values with samples colored by collection date.

on average every 2 weeks over 11 months from July 2021 to June 2022 (Fig. 1B) which was the period when the SARS-CoV-2 VOC lineages Delta and Omicron emerged. Viral load for each sample was quantified by real-time reverse transcriptase polymerase chain reaction (RT-qPCR) using four target primers (E, N1, ORF1ab, and S) (Supplementary Table 1). The most consistent results for viral load were obtained with the N1 primer (Fig. 1C-D).

The whole-genome amplicon sequencing data of wastewater samples was quantified by the breadth of coverage, which was calculated by mapping sequencing reads to the SARS-CoV-2 reference genome (GenBank accession # MN908947; RefSeq ID NC_045512.2). Of the 807 samples, 351 (43 %) had over 90 % coverage, 263 (33 %) had between 90 % and 50 % coverage, and 193 (24 %) had below 50 % coverage. The 614 samples with >50 % coverage passed the quality controls and were further used for downstream analysis. As expected, we observed that the breadth of coverage was highly correlated with the mean depth of coverage (Fig. 1C and E). The quality of the sequencing data was also consistent among catchments within states (Fig. 1C) despite the fact that different environmental factors and collection methods can directly impact SARS-CoV-2 RNA degradation in wastewater (Bertels et al., 2022) and viral genome recovery through high-throughput sequencing. These results supported the reliability of the sample processing performed in the study where the correlation of breadth of coverage with mean read depth and cycle threshold value observed are consistent across the collection period (Fig. 1E and F). Although the majority of samples with high viral RNA concentration via qPCR (corresponding to low cycle threshold values) showed high breadth of coverage in the Illumina sequencing, there was no clear correlation between the two parameters. In fact, some samples with low viral RNA concentration (high cycle threshold values) showed good breadth of coverage (Fig. 1D and F). This same observation has also been reported in previous studies (Crits-Christoph et al., 2021; Fontenele et al., 2021; Izquierdo-Lara et al., 2021).

3.2. Ordination analysis reveals the temporal shifts in VOC lineages prevalence

Unlike in clinically-based genomic epidemiology, it is not reasonable to use a consensus genome to represent the SARS-CoV-2 variants present in the wastewater as the samples will always contain a mixture of viral lineages. Ordination analysis, which has been widely implemented in metagenomic research, cannot be directly applied to SARS-CoV-2 whole genome sequencing data from wastewater because the canonical population dissimilarity measurement depends on knowledge of the component composition. To address this issue, we developed a tool called CoVdist to measure the pairwise distances between SARS-CoV-2 populations and estimate viral population diversity within wastewater samples (Fontenele et al., 2021) and between wastewater samples and SARS-CoV-2 lineages.

Using CoVdist, we performed ordination analysis on the wastewater samples (Fig. 2B). We included all the SARS-CoV-2 lineages from the Delta and Omicron VOC lineages (as assigned by pangolin v.4.1.2-pdata-1.13) (O'Toole et al., 2021), the two main VOC lineages circulating during the period of this study. The PCoA plot shows lineages from each VOC lineage cluster according to their respective clades in the global phylogenetic tree as described by Nextstrain (Hadfield et al., 2018) and highlights the observed temporal shift in the viral population. SARS-CoV-2 sequences in samples collected from July 2021 to early November 2021 clustered with Delta lineages, which were the most abundant circulating lineages based on clinical data. The majority of samples clustered more closely to the Delta lineages from clade 21J. The cluster of wastewater samples dominated by Delta lineages was significantly separated from those dominated by Omicron lineages along the first principal coordinate (PC1, accounting for 55.84 % of the total variance). The VOC lineage Omicron emerged in early December 2021 and had superseded Delta lineages by late December 2021 or early January 2022 at most locations. The population transition from Delta to Omicron lineages was also observed in the PCoA plot with samples from December and January clustering with the Omicron lineages from clade 21K (lineages BA.1 and descendants) and later samples (February/March/April 2022) clustering more closely to lineages in the clade 21L (BA.2, BA.5 and BA.4 and descendants). This shift once again mirrored what has been observed in clinical data where BA.1 derived lineages were more abundant in the initial Omicron wave of infection and were later displaced by BA.2 lineages. These results showed that the CoVdist viral diversity analysis can capture the temporal shifts in VOC lineages diversity in wastewater samples.

An analysis of sequences from samples per state demonstrated that specific timing of transitions between different VOC lineages differed by location. The shift from Delta to Omicron started in December 2021 for all states and only in January 2022 had Delta been replaced by Omicron. The timing of replacement of Omicron lineage BA.1 and descendants (21K clade) to BA.2 and descendants (21L clade) varies by state (Supplementary Fig. 1). Notably an earlier shift was observed in the states of Arizona, Washington, and Texas which occurred in February 2022, contrasting with the clinical data from these states where the transition occurred in March 2022. In all other states, the 21K to 21L transition started in March 2022, and it seemed to be finished by April 2022. The two exceptions are the states of New Mexico and Georgia where the complete replacement of 21K to 21L occurred only by May 2022, which in this case is a month later than what is observed in clinical data. Arizona was the state with the most collection sites (number of catchments = 17), which showed that the overall viral diversity can vary significantly within states and neighboring cities. This highlights the importance of wastewater surveillance at the neighborhood-level to identify changes in trends among VOC lineages with more refinement and inform public responses before the virus is disseminated city-wide.

Overall, the sample diversity analysis showed temporal changes in viral diversity within and between samples. When coupled with VOC lineages genetic diversity, it can also indicate the most prevalent VOC lineage per sample. This approach can be a valuable additional tool to observe global trends based on viral diversity and wastewater surveillance of viruses.

3.3. Defining mutations highlight transitions between SARS-CoV-2 VOC lineages

We further investigated the genetic information that supports the presence of major circulating VOC lineages (Delta and Omicron) in the wastewater by identifying the frequency of the defining mutations (SNVs, insertions, and deletions) associated with each VOC lineage. The frequency of defining SNVs per sample shows the same temporal trends that explain the displacement of VOC lineages as in the ordination analysis (Fig. 3). We were also able to detect the transition period when defining SNVs from both VOC lineages were detected.

The samples from Arizona were from 17 catchments in 5 cities (Mesa, Gilbert, Guadalupe, Tempe, and Scottsdale). The defining SNVs from Delta clades 21A and 21J were prevalent starting from July 2021 until December 2021 (Supplementary Fig. 2). In late December, there was a transition from Delta to Omicron during which mutations from both lineages appeared simultaneously. Those transition points were clear when we observed the defining mutations of each VOC lineage. As of January 2022, all catchments were dominated by defining mutations from the Omicron clade 21K (BA.1 and descendant lineages) except for one catchment in the city of Tempe (Tempe 03) that still presented high frequency of Delta-associated SNVs in early January 2022 (Supplementary Fig. 2). This result demonstrated that there could be differences in circulating lineages even at a neighborhood level, since all other catchments from Tempe had already shifted to Omicron lineages. In February 2022, we observed defining mutations from Delta but also from Omicron clade 21L in multiple catchments, demonstrating the utility of WBE in providing refined information on low frequency circulating lineages. During February and March, we observed the transition from Omicron clade 21K to Omicron clade 21L, with 21K becoming fully displaced by April 2022. This transition was only detectable in a subset of the catchments (Scottsdale, Tempe 01, 02, 04, 06, 08, and 09) (Supplementary Fig. 2).

There were only two states, Georgia and New Mexico, where Alpha-defining SNVs were still observed in samples from July 2021. However, they had already begun to be replaced by the Delta lineages as evidenced

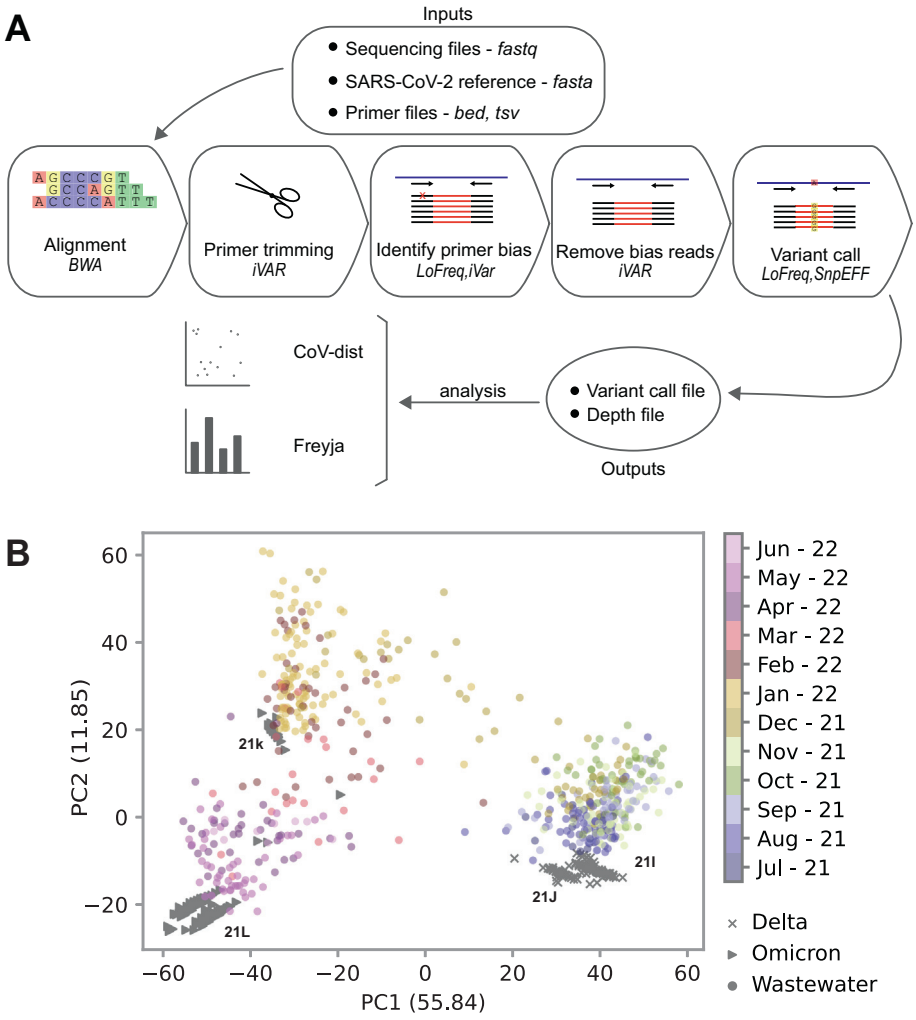


Fig. 2. Processing pipeline used to analyze the sequencing data for viral diversity analysis of all wastewater samples from this study. A. Description of the input files, steps, and output files for processing and analysis of wastewater samples from this study. B. Principal coordinate analysis (PCoA) including the samples that passed quality control from our study along with the pangolin lineages that are part of the VOC lineages Omicron (gray triangles) and Delta (gray X). The VOC lineages have been labeled based on the name used in the Nextstrain phylogenetic tree. The wastewater samples (circles) are color coded by collection date. The x and y axis represent the first and second principal coordinate, respectively, and the values in brackets represent the percentage of variations explained.

by the presence of defining mutations for both VOC lineages. (Supplementary Fig. 3). The replacement of defining mutations of Delta to Omicron clade 21K occurred by December 2021 in the wastewater samples from Georgia, Illinois, and New Mexico (Supplementary Fig. 3). However, in the states of Kentucky and New Mexico, we only observed the replacement in January 2022. Because we were lacking samples from December in the Kentucky locations, we cannot be certain if the replacement occurred before January 2022. Later, the transition from the Omicron lineages BA.1 (21K clade) to BA.2 lineages (clade 21L) occurred as early as January 2022 for the Illinois location and February 2022 for the Georgia and New Mexico locations in (Supplementary Fig. 3). The exact time of shift from 21K lineages to 21L is unclear for New Jersey due to missing sampling data, but the replacement was complete by April 2022. In the locations from the states of Oregon, Texas, and Washington the transition from Delta to Omicron started in December and in some locations, it lasted through January 2022 (Supplementary Fig. 4). Oregon and Washington samples from January 2022 had low variability in the VOC lineage abundance Delta/Omicron that we could not observe through viral diversity analysis (Supplementary Fig. 4). The further replacement of the Omicron lineages from clade 21K to 21L seemed to occur later in samples from the states of Oregon and Washington starting in March 2022 as opposed to February as seen in Texas and other states (Supplementary Fig. 4).

We observed the presence of multiple defining mutations of the Delta 21A clade but at a much lower allelic frequency, supporting the minor increase in Delta abundance for the locations of “Tempe 01”, “Americus”, “Carol Stream”, and “Des Moines 2” despite the previous replacement of Delta by Omicron (Fig. 3). This showed that Delta lineages were still present at these locations but were largely missed by the clinical data.

The Omicron lineages became prevalent after late December 2021 in the United States, but several Omicron defining mutations were detected in the wastewater samples collected from dates significantly earlier (Supplementary Figs. 2-4). The SNV C15240T was identified as early as August 2021 in Gilbert 1, Tempe 01, and Tempe 05; October 2021 in Tempe 04; and November 2021 in Tempe 02 and Scottsdale. The SNV C12880T was detected in Gilbert 2 in September 2021 and at Location 2 of Arizona in November 2021. The mutation S:Q493R was observed in the sample from Hoboken, New Jersey collected in September 2021 (Fig. 3) The defining insertion of Omicron (22,204:GAGCCAGAA) appeared in early December samples from Arizona before Omicron became prevalent. In fact, taking Arizona as an example, the earliest genomes isolated from clinical data deposited in GISAID belonging to the Omicron lineages date back to early March 2021. This indicates Omicron could have been circulating much earlier but at extremely low frequencies. The wastewater samples were able to capture the Omicron lineages that were circulating at a low frequency.

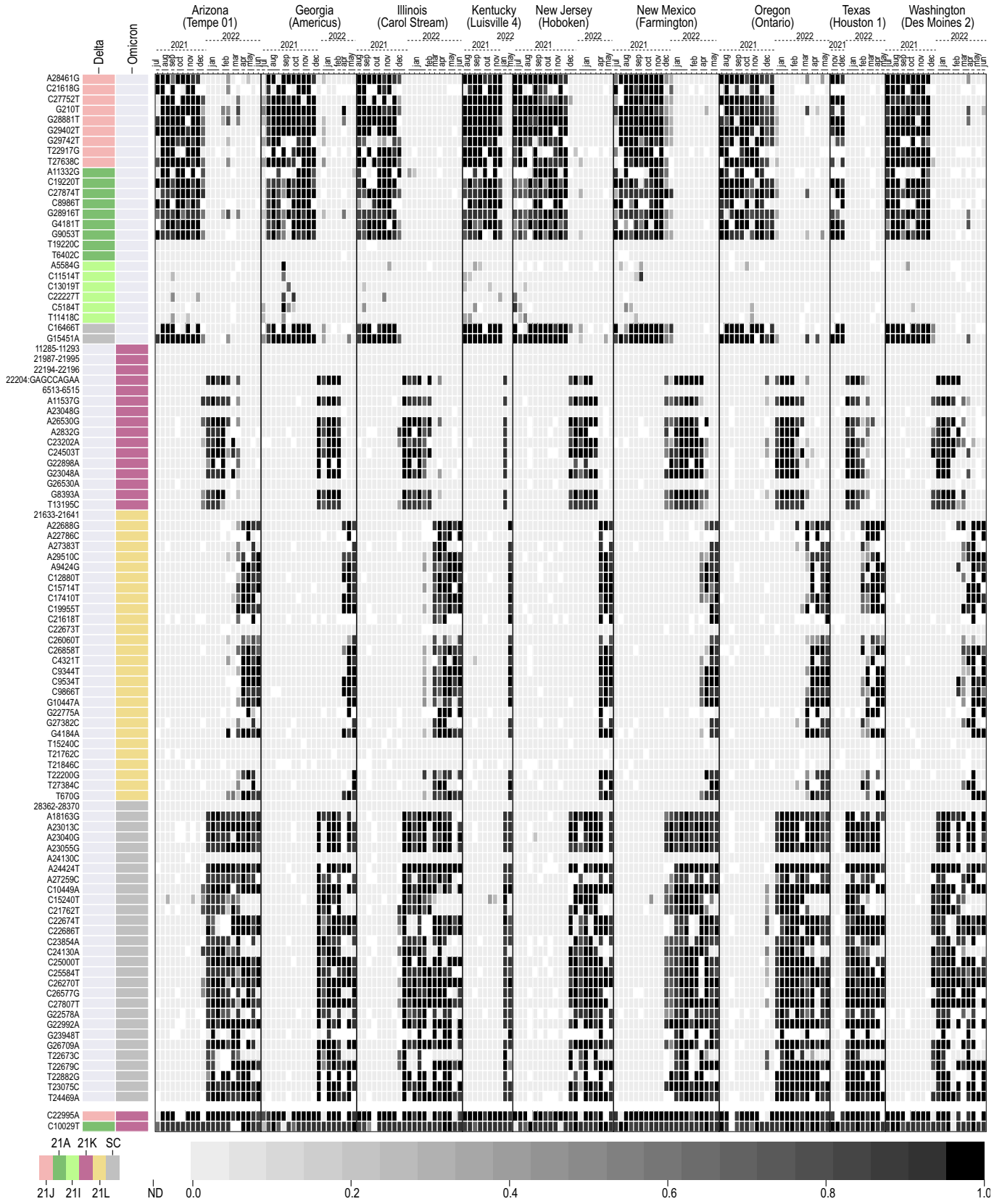


Fig. 3. Frequency of defining variants from Delta and Omicron lineages present in wastewater samples from one catchment in each state. The heatmap shows on the y-axis the variant that is associated with each VOC lineage. The first columns group those variants per VOC lineage which are color coded by the specific Nextstrain clade of the associated lineages. Variants shared by clades of the same VOC lineage are colored gray. At the top of the heatmap on the x-axis, each column represents a sample which is grouped by state and catchment and ordered by collection time. Each group of samples from the same catchment/state are also separated in boxes. The heatmap color shows the frequency of the defining variants that are present in each sample. If the variant position had no depth of coverage, the color is white.

The overall changes of the frequency of the defining mutations are consistent with the global VOC lineages trends, yet the specific time of the emergence of lineage defining mutations varied by geographical location, revealing regional variations in the transmission patterns of VOC lineages. In addition, the presence of Delta SNVs when Omicron was the most prevalent VOC lineage in clinical data and the early detection of Omicron defining mutations shows that wastewater sequencing can capture low frequency circulating lineages.

3.4. Lineage relative abundance variability at the state, city, and neighborhood level

The wastewater samples from the catchments revealed neighborhood and city-level variations of viral composition. We computed the relative abundance of SARS-CoV-2 lineages for each catchment with Freyja (Karthikeyan et al., 2022). Our results showed there were variations when comparing catchments at all geographical scales in the study. This variation could be attributed to the size of the population and city dynamic (residential or commercial setting) which influences population transit between locations and can lead to differences even between close neighborhoods.

Most of the observed variation occurred after January 2022 and the emergence of the Omicron lineages which includes the detection of Delta/Omicron recombinant sequences or recombinants of Omicron lineages. Based on clinical data, the initial wave of Omicron was mostly associated with the BA.1 lineage and descendants, and those lineages were subsequently displaced by BA.2 lineages and descendants (Fig. 4). However, we observed that depending on the catchment location of the wastewater samples, this transition period was also characterized by the presence of recombinant sequences between Omicron lineages, Delta lineages or other lineages, a more refined snapshot than what is observed by clinical data (Fig. 4 and Fig. 5) (Bolze et al., 2022; Focosi and Maggi, 2022; Lacek et al., 2022). Our results show that the presence of recombinant sequences is much higher than what has been documented by clinical data and that Omicron lineages' abundance varied significantly by catchment location.

The state of Arizona had the most catchments (n = 17) representing the cities of Gilbert, Guadalupe, Scottsdale, Tempe, and one undisclosed city in our study. The disclosed locations are in Maricopa County, part of the greater Phoenix metropolitan area, and some share geographical borders which means there should be a certain level of population exchange between cities. Unfortunately, we did not obtain consistent data across the collection period for all catchments and Tempe 03, Tempe 10, and Tempe 11 have very uneven data. Nonetheless, it is clear from the results that neighborhood scale surveillance shows significant variation in terms of VOC lineage abundance (Fig. 5). In some catchments, there are clearly different transition periods from BA.1 to BA.2 in which recombinant sequences are also detected. Interestingly we observe the emergence of the Omicron lineage BA.5 very early in March 2022 for the catchments Tempe 02 and Tempe 03 (Fig. 5). Even though some BA.5 genomes from clinical data in the state of Arizona were deposited in March 2022, the incidence of BA.5 in clinical data did not start rising until May 2022 (Fig. 4 and

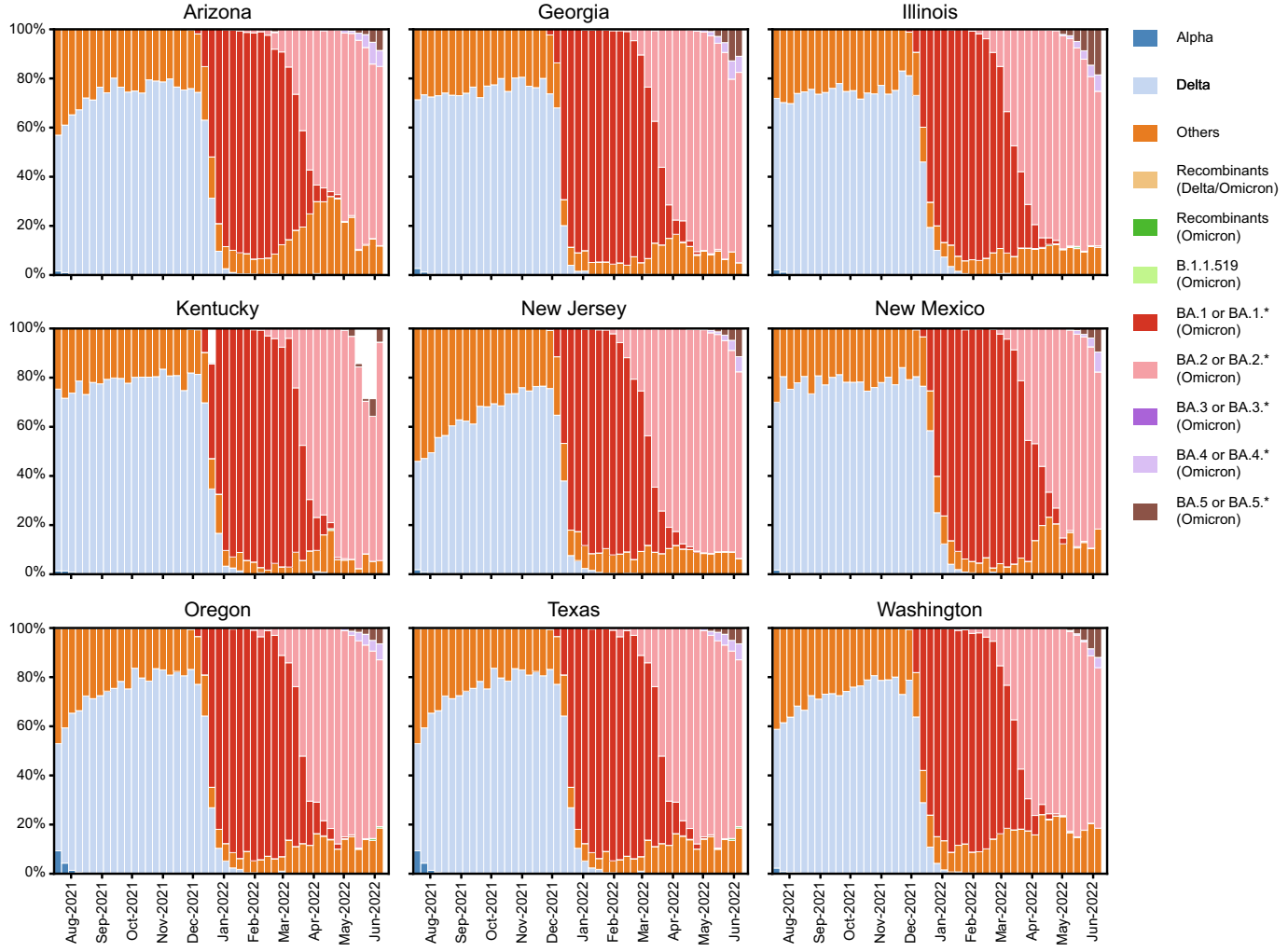


Fig. 4. Abundance of variants of concern based on clinical data for each state from this study. Bar plots showing the abundance of circulating variants of concern based on clinical data retrieved from GISAID using outbreak.info.

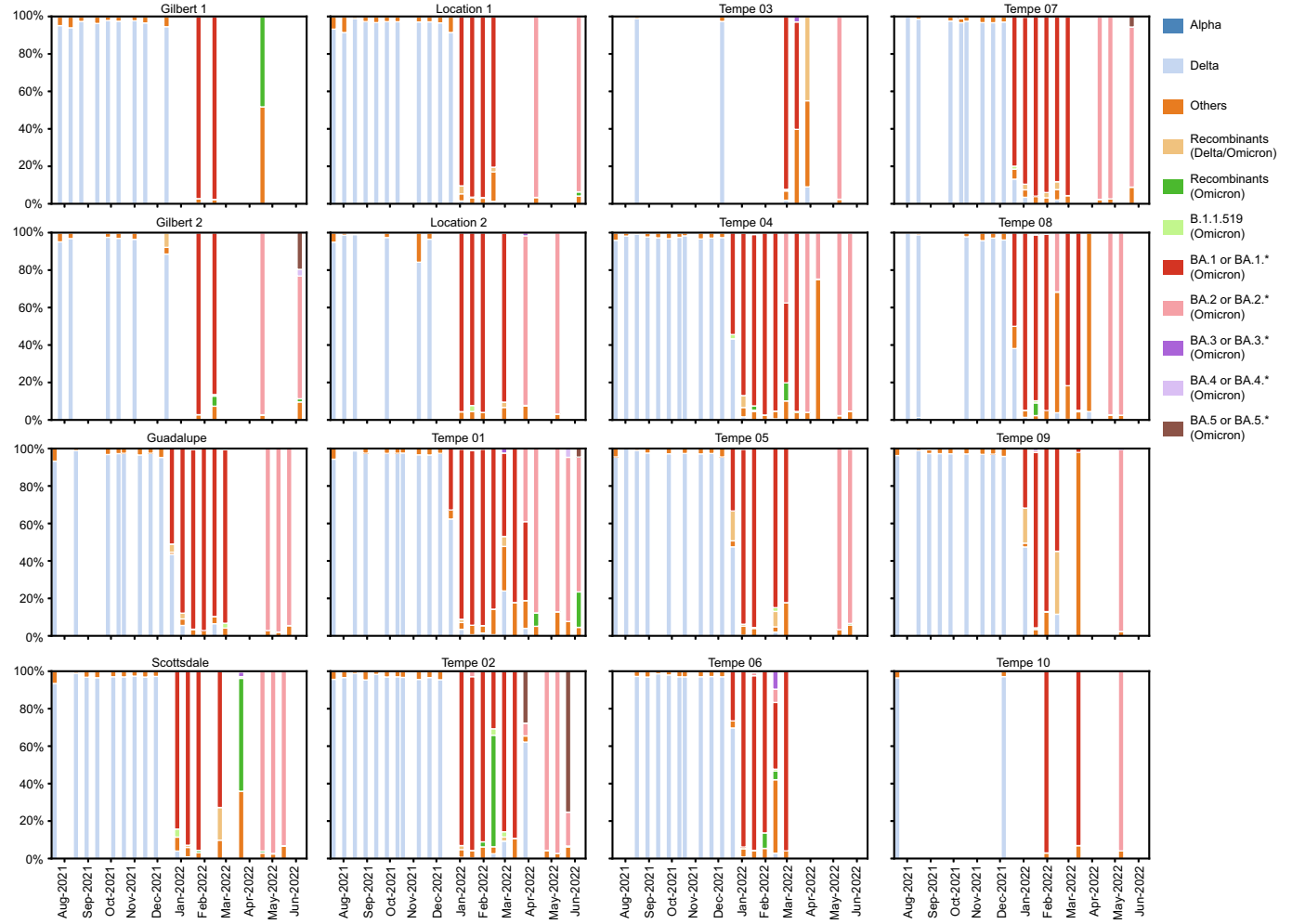


Fig. 5. Abundance of variants of concern per catchment in Arizona. Bar plots showing the abundance of circulating variants of concern (VOC) calculated using Freyja (y-axis) by collection time (x-axis). Bars are color-coded VOC lineages from Alpha, Delta, Omicron, and recombinant lineages. All other lineages are grouped into "other". The panel shows the catchments from the state of Arizona which include neighborhood and city-level collection sites.

Fig. 5). Our results show that this lineage was already circulating in Tempe in those two neighborhoods.

For the states of Georgia, Illinois, New Jersey, and New Mexico, there was only one catchment location for each (namely the cities of Americus, Carol Stream, Hoboken, and Farmington, respectively). Despite the bias in the sampling dates available for each location, we can observe the presence of Omicron recombinant lineages in the cities of Americus, Georgia and Farmington, New Mexico during a period of transition between the BA.1 lineages and BA.2 lineages (Supplementary Fig. 5). In Carol Stream, Illinois, we also observed the presence of lineage BA.4 starting in May 2022 (Supplementary Fig. 5) which corroborates what was seen for clinical data in the state of Illinois (Fig. 4). However, even though the BA.4 and BA.5 lineages had started to appear circulating in the states of Georgia, New Jersey, and New Mexico, those lineages were not detected in our collection cities.

In the state of Washington, we have catchments representing four cities and from those, the city of Sequim and Des Moines had two intra-city catchments. The overall abundance of VOC lineages trend is similar between the two, but there is a noticeable variation of Omicron lineage abundance between catchments of the same city which highlights once again the importance of neighborhood-level surveillance (Supplementary Fig. 5). We also report the emergence of Omicron lineages BA.4 and BA.5 in some but not all catchments, showing that the state-level clinical data surveillance does not reflect city-specific variation (Supplementary Fig. 5).

The three catchment locations from the state of Oregon are from different cities within different counties. Therefore, variability across catchments is expected. In Gresham and Dallas, we observed the presence of BA.4 and BA.5 in June 2022, as is observed in clinical data (Supplementary Fig. 5). On the contrary, the three catchments from Texas are all from within the city of Houston, and surprisingly we observed distinct differences in the samples from the end of May 2022 in which location Houston 3 seems to have an increase on Delta lineages even after Omicron has already seemingly replaced Delta lineages (Supplementary Fig. 5). Due to the lack of consistent data from the following months across the three Houston locations, it is difficult to conclude if there was any other abundance variation occurring. Lastly, we observed no variation across the six collection sites of Louisville in Kentucky and, even though we are missing data from some months, the data appears to be consistent between catchments.

4. Discussion

Here, we demonstrate that targeted high-throughput sequencing of wastewater can provide relevant information on viral diversity and circulating variants of SARS-CoV-2 to help inform public health responses. The integrity of the viral genome can be affected by added chemicals, shifts in temperature, pH, and many other environmental factors that will influence the recovery of the genome through sequencing (Bertels et al., 2022). In addition, the viral genomic information present in wastewater represents all

infected individuals that contribute to the catchment in the area which creates a challenge for the analysis of the sequencing results. Tools to analyze data from WBE are underdeveloped. Therefore, we developed a tool called CoVdist to assess viral diversity using ordination analysis. This type of analysis has been widely applied to traditional metagenomics but has not been applied to whole-genome amplicon sequenced data including WBE. CoVdist allowed us to observe temporal trends in wastewater viral diversity in the context of the genetic composition of SARS-CoV-2 lineages, supporting catchment-level analysis.

We compared the efficacy of WBE and clinical genomic epidemiology at the state level (1st administrative division) due to a lack of city-specific location metadata in GISAID (Elbe and Buckland-Merrett, 2017; Khare et al., 2021; Shu and McCauley, 2017). Although this is an impartial comparison, this study shows that examining only clinical data at the state level does not reflect the actual diversity of lineages circulating in each city. Although other studies showed a stronger correlation between the diversity of lineages identified in wastewater and clinical data for some cities (Agrawal et al., 2022; Baaijens et al., 2022), this could be affected by higher rates of testing in those locations. Additionally, these studies were done at the city scale, but their conclusions may have been different if a neighborhood scale analysis was performed like in our study. Nevertheless, our results show how relevant WBE can be for providing a cost-effective surveillance tool to inform public health responses, especially for cities that might not have the capacity for clinical testing and sequencing. The detection of much more variation on VOC lineage abundances in the locations analyzed can be related to low clinical sequencing in the area, but it can also be because wastewater samples provide genetic information on non-symptomatic individuals which are mostly overlooked by clinical sequencing.

It is important to highlight that movement of populations across cities and neighborhoods may also influence the variation of lineage abundance. In fact, in the catchments from adjacent parts of the Phoenix greater area where mobility is expected to be high, we observed the most variation across neighborhoods. In contrast, a large-scale study done in Austria did not identify strong correlation between mobility and SARS-CoV-2 genetic diversity (Amman et al., 2022) indicating that this effect may vary in different cities. This city-specific trend can also be noticed by our data from Louisville which although representing city and neighborhood level do not show much variation. The same results have also been reported by another study done in Louisville (Rouchka et al., 2021). Another important factor that is not well studied is the difference in extended shedding from infection by different lineages, which may influence the abundance in the wastewater but may not be reflected in clinical data. These factors, along with other demographic and socioeconomic differences between regions, likely have a significant impact on the regional differences seen in our study. Nevertheless, it is very likely that the variance in wastewater is influenced by lineages circulating that have not been observed by clinical data.

As SARS-CoV-2 continues to spread and the virus continues to evolve, it is likely that viral population dynamics will continue to change and that new lineages will emerge. As less emphasis is placed on clinical testing, WBE will be a valuable approach for continuing to monitor the population dynamics and spread of SARS-CoV-2 at different regional scales. The tools and approaches developed in this study advance the WBE field generally, providing new ways to analyze WBE samples. Additionally, these methods are not specific to SARS-CoV-2 and could be easily adapted to monitor future outbreaks of viruses.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.162862.

Code availability

The SARS-CoV-2 metagenomic data analysis pipeline is available in a GitHub repository (https://github.com/nlm-irp-jianglab/bioinfo-wwbe). CoVdist is available in a GitHub repository (https://github.com/nlm-irp-jianglab/CoVdist).

CRediT authorship contribution statement

Rafaela S. Fontenele: Methodology, Software, Formal analysis, Data curation, Visualization, Writing – original draft. **Yiyan Yang:** Methodology, Software, Visualization, Writing – review & editing. **Erin M. Driver:** Software, Resources, Writing – review & editing. **Arjun Magge:** Software, Writing – review & editing. **Simona Kraberger:** Resources, Writing – review & editing. **Joy M. Custer:** Resources, Writing – review & editing. **Keith Dufault-Thompson:** Data curation, Writing – review & editing. **Erin Cox:** Resources. **Melanie Engstrom Newell:** Resources. **Arvind Varsani:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing. **Rolf U. Halden:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing. **Matthew Scotch:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing. **Xiaofang Jiang:** Methodology, Software, Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Data availability

Raw sequencing data were deposited to NCBI with the BioProject Accession number PRJNA847239. Metadata associated with the wastewater samples from this study is available in Supplementary Table 1.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: RUH and EMD are co-founders of AquaVitas, LLC, Phoenix, Arizona, United States, an Arizona State University startup company providing commercial services in wastewater-based epidemiology. RUH also is the founder of OneWaterOneHealth, a nonprofit project of the Arizona State University Foundation.

Acknowledgements

This work was supported in part by grants from the National Institutes of Health under Award Number U01LM013129 under the RADx-rad initiative for emergency response to COVID-19 to RUH, MS, and AV. RSF, YY, KD and XJ are supported by the Intramural Research Program of the NIH, National Library of Medicine. This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov). Samples were acquired with partial support from an award by the J.M. Kaplan Fund to RUH: OneWaterOneHealth nonprofit project 30009070 of the Arizona State University Foundation. The authors gratefully acknowledge the originating and submitting laboratories who contributed sequences to GISAID (www.gisaid.org). The authors would like to thank Tyler Perleberg, Allan Yanez, Izabella Block, Anumitha Aravindan, Ayesha Babbrah and Erin Clancy from the ASU Biodesign Center for Environmental Health Engineering for their support. This work would not be possible without the participation of the municipalities, and we are deeply appreciative.

References

Agrawal, S., Orschler, L., Schubert, S., Zachmann, K., Heijnen, L., Tavazzi, S., Gawlik, B.M., de Graaf, M., Medema, G., Lackner, S., 2022. Prevalence and circulation patterns of SARS-CoV-2 variants in european sewage mirror clinical data of 54 european cities. Water Res. 214, 118162.

Aksamentov, I., Roemer, C., Hodcroft, E., Neher, R., 2021. Nextclade: clade assignment, mutation calling and quality control for viral genomes. J. Open Source Softw. 6, 3773.

Amman, F., Markt, R., Endler, L., Hupfauf, S., Agerer, B., Schedl, A., Richter, L., Zechmeister, M., Bicher, M., Heiler, G., Triska, P., Thornton, M., Penz, T., Senekowitsch, M., Laine, J., Keszei, Z., Klimek, P., Nägele, F., Mayr, M., Daleiden, B., Steinlechner, M., Niederstätter, H., Heidinger, P., Rauch, W., Scheffknecht, C., Vogl, G., Weichlinger, G., Wagner, A.O., Slipko, K., Masseron, A., Radu, E., Allerberger, F., Popper, N., Bock, C., Schmid, D., Oberacher, H., Kreuzinger, N., Insam, H., Bergthaler, A., 2022. Viral variant-resolved wastewater surveillance of SARS-CoV-2 at national scale. Nat. Biotechnol. 40, 1814–1822.

Baaijens, J.A., Zulli, A., Ott, I.M., Nika, I., van der Lugt, M.J., Petrone, M.E., Alpert, T., Fauver, J.R., Kalinich, C.C., Vogels, C.B.F., Breban, M.I., Duvallet, C., McElroy, K.A., Ghaeli, N.,

Imakaev, M., Mckenzie-Bennett, M.F., Robison, K., Plocik, A., Schilling, R., Pierson, M., Littlefield, R., Spencer, M.L., Simen, B.B., Yale SARS-CoV-2 Genomic Surveillance Initiative, Hanage, W.P., Grubaugh, N.D., Peccia, J., Baym, M., 2022. Lineage abundance estimation for SARS-CoV-2 in wastewater using transcriptome quantification techniques. Genome Biol. 23, 236.

Bertels, X., Demeyer, P., Van den Bogaert, S., Boogaerts, T., van Nuijs, A.L.N., Delputte, P., Lahousse, L., 2022. Factors influencing SARS-CoV-2 RNA concentrations in wastewater up to the sampling stage: a systematic review. Sci. Total Environ. 820, 153290.

Bolze, A., Basler, T., White, S., Dei Rossi, A., Wyman, D., Dai, H., Roychoudhury, P., Greninger, A.L., Hayashibara, K., Beatty, M., Shah, S., Stous, S., McCrone, J.T., Kil, E., Cassens, T., Tsan, K., Nguyen, J., Ramirez, J., Carter, S., Cirulli, E.T., Schiabor Barrett, K., Washington, N.L., Belda-Ferre, P., Jacobs, S., Sandoval, E., Becker, D., Lu, J.T., Isaksson, M., Lee, W., Luo, S., 2022. Evidence for SARS-CoV-2 Delta and omicron co-infections and recombination. Med (N Y) 3, 848–859 e4.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6, 80–92.

Corman, V.M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D.K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M.L., Mulders, D.G., Haagmans, B.L., van der Veer, B., van den Brink, S., Wijsman, L., Goderski, G., Romette, J.-L., Ellis, J., Zambon, M., Peiris, M., Goossens, H., Reusken, C., Koopmans, M.P., Drosten, C., 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill. 25. https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045.

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., Hetzel, J., Spitzer, S., Koble, J., Tan, A., Hyde, F., Schroth, G., Kuersten, S., Banfield, J.F., Nelson, K.L., 2021. Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 variants. MBio 12. https://doi.org/10.1128/mBio.02703-20.

Elbe, S., Buckland-Merrett, G., 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. Global Chall. 1, 33–46.

Ellmen, I., Lynch, M.D.J., Nash, D., Cheng, J., Nissimov, J.I., Charles, T.C., 2021. Alcov: Estimating Variant of Concern Abundance From SARS-CoV-2 Wastewater Sequencing Data. medRxiv 2021–2006.

Focosi, D., Maggi, F., 2022. Recombination in coronaviruses, with a focus on SARS-CoV-2. Viruses 14. https://doi.org/10.3390/v14061239.

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., Holmes, W.K., Deitrick, S., Brown, P., Duty, D., Smith, T., Bhatnagar, A., Yeager II, R.A., Holm, R.H., von Reitzenstein, N.H., Wheeler, E., Dixon, K., Constantine, T., Wilson, M.A., Lim, E.S., Jiang, X., Halden, R.U., Scotch, M., Varsani, A., 2021. High-throughput sequencing of SARS-CoV-2 in wastewater provides insights into circulating variants. Water Res. 205, 117710.

Gangavarapu, K., Latif, A.A., Mullen, J.L., Alkuzweny, M., Hufbauer, E., Tsueng, G., Haag, E., Zeller, M., Aceves, C.M., Zaiets, K., Cano, M., Zhou, X., Qian, Z., Sattler, R., Matteson, N.L., Levy, J.I., Lee, R.T.C., Freitas, L., Maurer-Stroh, S., GISAID Core and Curation Team, Suchard, M.A., Wu, C., Su, A.I., Andersen, K.G., Hughes, L.D., 2023. Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. Nat. Methods https://doi.org/10.1038/s41592-023-01769-3.

Gregory, D.A., Wieberg, C.G., Wenzel, J., Lin, C.-H., Johnson, M.C., 2021. Monitoring SARS-CoV-2 populations in wastewater by amplicon sequencing and using the novel program SAM refiner. Viruses 13. https://doi.org/10.3390/v13081647.

Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., Tan, A.L., Paul, L.M., Brackney, D.E., Grewal, S., Gurfield, N., Van Rompay, K.K.A., Isern, S., Michael, S.F., Coffey, L.L., Loman, N.J., Andersen, K.G., 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 20, 8.

Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R.A., 2018. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34, 4121–4123.

Izquierdo-Lara, R., Elsinga, G., Heijnen, L., Oude Munnink, B.B., Schapendonk, C.M.E., Nieuwenhuijse, D., Kon, M., Lu, L., Aarestrup, F.M., Lycett, S., Medema, G., Koopmans, M.P.G., de Graaf, M., 2021. Monitoring SARS-CoV-2 circulation and diversity through community wastewater sequencing, the Netherlands and Belgium. Emerg. Infect. Dis. https://doi.org/10.3201/eid2705.204410.

Jahn, K., Dreifuss, D., Topolsky, I., Kull, A., Ganesanandamoorthy, P., Fernandez-Cassi, X., Bänziger, C., Devaux, A.J., Stachler, E., Caduff, L., Cariti, F., Corzón, A.T., Fuhrmann, L., Chen, C., Jablonski, K.P., Nadeau, S., Feldkamp, M., Beisel, C., Aquino, C., Stadler, T., Ort, C., Kohn, T., Julian, T.R., Beerenwinkel, N., 2022. Early detection and surveillance of SARS-CoV-2 genomic variants in wastewater using COJAC. Nat. Microbiol. 7, 1151–1160.

Jones, D.L., Baluja, M.Q., Graham, D.W., Corbishley, A., McDonald, J.E., Malham, S.K., Hillary, L.S., Connor, T.R., Gaze, W.H., Moura, I.B., Wilcox, M.H., Farkas, K., 2020. Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19. Sci. Total Environ. 749, 141364.

Karthikeyan, S., Levy, J.I., De Hoff, P., Humphrey, G., Birmingham, A., Jepsen, K., Farmer, S., Tubb, H.M., Valles, T., Tribelhorn, C.E., Tsai, R., Aigner, S., Sathe, S., Moshiri, N., Henson, B., Mark, A.M., Hakim, A., Baer, N.A., Barber, T., Belda-Ferre, P., Chacón, M., Cheung, W., Cresini, E.S., Eisner, E.R., Lastrella, A.L., Lawrence, E.S., Marotz, C.A., Ngo, T.T., Ostrander, T., Plascencia, A., Salido, R.A., Seaver, P., Smoot, E.W., McDonald, D., Neuhard, R.M., Scioscia, A.L., Satterlund, A.M., Simmons, E.H., Abelman, D.B., Brenner, D., Bruner, J.C., Buckley, A., Ellison, M., Gattas, J., Gonias, S.L., Hale, M., Hawkins, F., Ikeda, L., Jhaveri, H., Johnson, T., Kellen, V., Kremer, B., Matthews, G., McLawhon, R.W., Ouillet, P., Park, D., Pradenas, A., Reed, S., Riggs, L., Sanders, A., Sollenberger, B., Song, A., White, B., Winbush, T., Aceves, C.M., Anderson, C., Gangavarapu, K., Hufbauer, E., Kurzban, E., Lee, J., Matteson, N.L., Parker, E., Perkins, S.A., Ramesh, K.S., Robles-Sikisaka, R., Schwab, M.A., Spencer, E., Wohl, S., Nicholson, L., McHardy, I.H., Dimmock, D.P., Hobbs, C.A., Bakhtar, O., Harding, A., Mendoza, A., Bolze, A., Becker, D., Cirulli, E.T., Isaksson, M., Schiabor Barrett, K.M., Washington, N.L., Malone, J.D., Schafer, A.M., Gurfield, N., Stous, S., Fielding-Miller, R., Garfein, R.S., Gaines, T., Anderson, C., Martin, N.K., Schooley, R., Austin, B., MacCannell, D.R., Kingsmore, S.F., Lee, W., Shah, S., McDonald, E., Yu, A.T., Zeller, M., Fisch, K.M., Longhurst, C., Maysent, P., Pride, D., Khosla, P.K., Laurent, L.C., Yeo, G.W., Andersen, K.G., Knight, R., 2022. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. Nature 609, 101–108.

Khare, S., Gurry, C., Freitas, L., Schultz, M.B., Bach, G., Diallo, A., Akite, N., Ho, J., Lee, R.T., Yeo, W., Curation Team, G.C, Maurer-Stroh, S., 2021. GISAID's role in pandemic response. China CDC Wkly. 3, 1049–1051.

Lacek, K.A., Rambo-Martin, B.L., Batra, D., Zheng, X.-Y., Sakaguchi, H., Peacock, T., Keller, M., Wilson, M.M., Sheth, M., Davis, M.L., Borroughs, M., Gerhart, J., Hassell, N., Shepard, S.S., Cook, P.W., Lee, J., Wentworth, D.E., Barnes, J.R., Kondor, R., Paden, C.R., 2022. Identification of a Novel SARS-CoV-2 Delta-Omicron Recombinant Virus in the United States. bioRxiv https://doi.org/10.1101/2022.03.19.484981.

Lanfear, R., Mansfield, R., 2020. A Global Phylogeny of SARS-CoV-2 Sequences From GISAID. Zenodo https://doi.org/10.5281/ZENODO.3958883.

Lavania, M., Joshi, M.S., Ranshing, S.S., Potdar, V.A., Shinde, M., Chavan, N., Jadhav, S.M., Sarkale, P., Mohandas, S., Sawant, P.M., Tikute, S., Padbidri, V., Patwardhan, S., Kate, R., 2022. Prolonged shedding of SARS-CoV-2 in feces of COVID-19 positive patients: trends in genomic variation in first and second wave. Front. Med. 9, 835168.

Layton, B.A., Kaya, D., Kelly, C., Williamson, K.J., Alegre, D., Bachhuber, S.M., Banwarth, P.G., Bethel, J.W., Carter, K., Dalziel, B.D., Dasenko, M., Geniza, M., George, A., Girard, A.-M., Haggerty, R., Higley, K.A., Hynes, D.M., Lubchenco, J., McLaughlin, K.R., Nieto, F.J., Noakes, A., Peterson, M., Piemonti, A.D., Sanders, J.L., Tyler, B.M., Radniecki, T.S., 2022. Evaluation of a wastewater-based epidemiological approach to estimate the prevalence of SARS-CoV-2 infections and the detection of viral variants in disparate Oregon communities at city and neighborhood scales. Environ. Health Perspect. 130, 67010.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25, 1754–1760.

Mantilla-Calderon, D., Huang, K.(Kevin), Li, A., Chibwe, K., Yu, X., Ye, Y., Liu, L., Ling, F., 2022. 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. 8, 1391–1407.

McBroome, J., Thornlow, B., Hinrichs, A.S., Kramer, A., De Maio, N., Goldman, N., Haussler, D., Corbett-Detig, R., Turakhia, Y., 2021. A daily-updated database and tools for comprehensive SARS-CoV-2 mutation-annotated trees. Mol. Biol. Evol. 38, 5819–5824.

O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J.T., Colquhoun, R., Ruis, C., Abu-Dahab, K., Taylor, B., Yeats, C., du Plessis, L., Maloney, D., Medd, N., Attwood, S.W., Aanensen, D.M., Holmes, E.C., Pybus, O.G., Rambaut, A., 2021. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol. 7, veab064.

Park, S.-K., Lee, C.-W., Park, D.-I., Woo, H.-Y., Cheong, H.S., Shin, H.C., Ahn, K., Kwon, M.-J., Joo, E.-J., 2021. Detection of SARS-CoV-2 in fecal samples from patients with asymptomatic and mild COVID-19 in Korea. Clin. Gastroenterol. Hepatol. 19, 1387–1394 e2.

Pechlivanis, N., Tsagiopoulou, M., Maniou, M.C., Togkousidis, A., Mouchtaropoulou, E., Chassalevris, T., Chaintoutis, S.C., Petala, M., Kostoglou, M., Karapantsios, T., Laidou, S., Vlachonikola, E., Chatzidimitriou, A., Papadopoulos, A., Papaioannou, N., Dovas, C.I., Argiriou, A., Psomopoulos, F., et al., 2022. Detecting SARS-CoV-2 lineages and mutational load in municipal wastewater and a use-case in the metropolitan area of Thessaloniki, Greece. Sci. Rep 12, 2659.

Pérez-Cataluña, A., Chiner-Oms, Á., Cuevas-Ferrando, E., Díaz-Reolid, A., Falcó, I., Randazzo, W., Girón-Guzmán, I., Allende, A., Bracho, M.A., Comas, I., Sánchez, G., 2022. Spatial and temporal distribution of SARS-CoV-2 diversity circulating in wastewater. Water Res. 211, 118007.

Rouchka, E.C., Chariker, J.H., Saurabh, K., Waigel, S., Zacharias, W., Zhang, M., Talley, D., Santisteban, I., Puccio, M., Moyer, S., Holm, R.H., Yeager, R.A., Sokoloski, K.J., Fuqua, J., Bhatnagar, A., Smith, T., 2021. The rapid assessment of aggregated wastewater samples for genomic surveillance of SARS-CoV-2 on a City-wide scale. Pathogens 10. https://doi.org/10.3390/pathogens10101271.

Shu, Y., McCauley, J., 2017. GISAID: global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 22. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494.

Sutton, M., Radniecki, T.S., Kaya, D., Alegre, D., Geniza, M., Girard, A.-M., Carter, K., Dasenko, M., Sanders, J.L., Cieslak, P.R., Kelly, C., Tyler, B.M., 2022. Detection of SARS-CoV-2 B.1.351 (Beta) variant through wastewater surveillance before case detection in a community, Oregon, USA. Emerg. Infect. Dis. 28, 1101–1109.

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

Turakhia, Y., Thornlow, B., Hinrichs, A.S., De Maio, N., Gozashti, L., Lanfear, R., Haussler, D., Corbett-Detig, R., 2021. Ultrafast sample placement on existing tRees (USHER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. Nat. Genet. 53, 809–816.

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

Wilm, A., Aw, P.P.K., Bertrand, D., Yeo, G.H.T., Ong, S.H., Wong, C.H., Khor, C.C., Petric, R., Hibberd, M.L., Nagarajan, N., 2012. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 40, 11189–11201.

Xiao, F., Tang, M., Zheng, X., Liu, Y., Li, X., Shan, H., 2020. Evidence for gastrointestinal infection of SARS-CoV-2. Gastroenterology 158, 1831–1833. https://doi.org/10.1053/j.gastro.2020.02.055 e3.

Xing, Y.-H., Ni, W., Wu, Q., Li, W.-J., Li, G.-J., Wang, W.-D., Tong, J.-N., Song, X.-F., Wing-Kin Wong, G., Xing, Q.-S., 2020. Prolonged viral shedding in feces of pediatric patients with coronavirus disease 2019. J. Microbiol. Immunol. Infect. 53, 473–480.

Yue, J.C., Clayton, M.K., 2005. A similarity measure based on species proportions. Commun. Stat. Theory Methods 34, 2123–2131.