medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

Wastewater genomic surveillance captures early detection of Omicron in Utah

Pooja Gupta[1]*, Stefan Liao[1], Maleea Ezekiel[1], Nicolle Novak[1], Alessandro Rossi[1], Nathan LaCross[1], Kelly Oakeson[1], and Andreas Rohrwasser[1]

[1] Utah Department of Health and Human Services, Salt Lake City, UT 84129, USA

*Corresponding author:

Pooja Gupta, PhD

Utah Public Health Laboratory, Utah Department of Health and Human Services,

4431 South 2700 West, Salt Lake City, Utah 84129

Phone: 706-308-3189

Email: pgupta@utah.gov

Abstract

Wastewater-based epidemiology (WBE) has emerged as a powerful public health tool to trace new outbreaks, detect trends in infection and provide an early warning of COVID-19 community spread. Here, we investigated the spread of SARS-CoV-2 infections across Utah by characterizing lineages and mutations detected in wastewater samples. We sequenced over 1,200 samples from 32 sewersheds collected between November 2021 and March 2022. Wastewater sequencing confirmed the presence of Omicron (B.1.1.529) in Utah in samples collected on November 19, 2021, up to seven days before its corresponding detection via clinical sequencing. We observed an increase in prevalence of several Omicron-defining mutations across sewersheds from December to January. Proportion of Omicron increased to ~58% by January 4th 2022 and completely displaced Delta by February 7th, 2022. Our study shows that WBE is an excellent complementary tool to SARS-CoV-2 clinical surveillance and can guide public health action and policy decisions.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

Keywords

SARS-CoV-2, Omicron, wastewater surveillance, whole genome sequencing, Utah, public health

Introduction

Since the beginning of the COVID-19 pandemic, the world has witnessed the evolution of the SARS-CoV-2 virus in real time, with the emergence of new variants and the rapid spread of COVID-19 affecting millions of lives. So far, five variants of concern (VOCs) have been verified by the World Health Organization (WHO): Alpha, Beta, Gamma, Delta and Omicron. Globally, many countries have now lifted COVID-related restrictions, limited vaccine mandates and transitioned to normality. Thus, there remains an increasing concern over the emergence of novel SARS-CoV-2 variants with increased virulence, transmissibility and immune escape potential affecting the efficacy of currently available vaccines (1,2). Over the course of the pandemic, COVID-19 surveillance strategies have relied on testing of symptomatic individuals and their contacts to slow the spread of the pandemic. However, symptoms seem to be less severe in the ongoing Omicron wave, with many affected individuals remaining asymptomatic and often not seeking clinical testing, thereby rendering traditional surveillance systems less reliable. Furthermore, increased adoption of at-home testing results in systematic underreporting and less efficient surveillance. Thus, as the pandemic shifts, relying solely on clinical testing may be insufficient for timely detection of outbreaks and emerging variants. To circumvent these challenges, there is an urgent need to incorporate alternative testing strategies to detect low, circulating levels of SARS-CoV-2 variants.

Wastewater surveillance of SARS-CoV-2 is one such approach that can be used to track transmission and spread of new and existing variants across communities (3–7). Although, wastewater-based epidemiology (WBE) has traditionally been used for surveillance of a variety

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

of targets, including poliovirus (8), antimicrobial resistance genes (9) and monitoring illicit drug-use (10), it has recently gained traction as a promising epidemiological tool to track SARS-CoV-2 infections. The Centers for Disease Control and Prevention (CDC) has developed a national database - the National Wastewater Surveillance System (NWSS) - in late 2020 to monitor the spread of COVID-19 in the United States (CDC, 2020). WBE of SARS-CoV-2 complements existing clinical surveillance as it can capture both symptomatic and asymptomatic cases. As SARS-CoV-2 is known to be shed in human feces post-infection, usually well before symptoms appear, the viral loads in sewersheds provides a good indicator of SARS-CoV-2 case burden and transmission. Previous studies have found strong correlation between SARS-CoV-2 concentrations in wastewater and reported COVID-19 cases (11–14).

WBE of SARS-CoV-2 is useful in monitoring infections when clinical testing is limited, expensive or overwhelmed due to case surges, especially during the onset of a new variant. It is also not subject to biases related to health care seeking behavior and willingness to be tested. Previous work has shown that WBE could be essential to conducting long-term, population-wide SARS-CoV-2 surveillance in a rapid and cost-effective manner (15). Importantly, WBE can provide an early warning of rising COVID-19 infections in communities (16,17) and could be an important tool for guiding public health interventions in combating the ongoing COVID-19 pandemic.

Despite the advantages of WBE for SARS-CoV-2 surveillance, fewer studies implement genomic analysis of wastewater samples. A wastewater sample represents pooled genomic content from multiple individuals and presents a mix of different viral lineages. Additionally, wastewater samples are often degraded and may contain fragmented viral genomes (18). Thus, it is often challenging to confidently assign lineages in a wastewater sample and assess viral

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

genetic diversity. Consequently, early research efforts have focused mainly on measuring virus titers, usually via qPCR, to track trends in COVID-19 infection (3,12,13) or digital droplet PCR (ddPCR) to characterize variants of interest (VOI) and VOCs (19,20). However, PCR-based approaches are limited in their application due to mutations being shared across VOI/VOCs and ongoing evolution of SARS-CoV-2 virus leading to the occurrence of novel mutations. It also requires continual updating and validation of existing methods to avoid primer dropouts. As ddPCR targets signature mutations, it may miss amplification of other possible variants which may be epidemiologically relevant. To circumvent these issues, recent studies have adapted the use of next generation sequencing approaches to comprehensively scan the entire SARS-CoV-2 genome for potential mutations (21–24). This enables detection of unique and functionally important mutations and VOC-defining clusters of mutations. Novel computational approaches are also being developed to parse the different viral lineages in wastewater samples (21,25,26). Genomic sequencing of SARS-CoV-2 in wastewater offers a universal and powerful tool to assess the prevalence of circulating SARS-CoV-2 variants and help detect new emerging variants in a population.

In this study, we sequenced wastewater samples collected from 32 sewersheds across Utah. Our goal was to track SARS-CoV-2 spread by characterizing lineages and mutations detected in wastewater communities. We also examined whether lineages detected in wastewater were comparable to lineage diversity observed in clinical samples. We collaborated with the Centre for Genomic Pathogen Surveillance to develop an interactive dashboard to visualize our wastewater sequencing data and inform local public health investigations. This internal dashboard provides an intuitive summary of wastewater sequencing results to our state

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

epidemiologists and helps in keeping track of various SARS-CoV-2 mutations and associated lineages over time and across various sampling sites.

Methods

Sample collection, processing and qPCR testing

Sewage samples were collected twice per week at 32 wastewater facilities between November 2021 and March 2022 (n=1,235; Figure 1; Appendix Table 1). 20mL of raw influent wastewater sample was processed following the steps outlined in the Promega Maxwell® RSC Enviro Total Nucleic Acid Kit (Promega Corp.). Sample collection and processing steps are detailed in Appendix methods. We conducted RT-qPCR using TaqPath mastermix and the primers and probes contained within the 2019-nCoV CDC EUA kit (27) to detect the presence of the targeted gene sequences(N1 and N2) on a ThermoScientific ABI 7500 Fast instrument. We quantified SARS-CoV-2 RNA against a 5-point standard curve made using serial dilutions of the CDC EUA 2019-nCoV positive control with N1 and N2 analyzed in combination.

Sequencing and Bioinformatics analysis

We sequenced the SARS-CoV-2 genome using amplicon-based sequencing following the Illumina COVIDSeq™ RUO protocol (Illumina, Inc. San Diego, CA, USA). Library preparation was automated via Tecan liquid handlers and pooled, amplified libraries were loaded onto the S4-flow cell following NovaSeq workflow as per manufacturer's instructions (Illumina Inc). Dual indexed single-end sequencing with a 75bp read length was performed on the NovaSeq 6000 platform.

After sequencing, we generated FASTQ files from the raw data using the DRAGEN COVIDSeq Test Pipeline (Illumina Inc.) on the Illumina DRAGEN Bio-IT platform.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

We used viralrecon (v.2.4.1) to analyze the sequencing data and detect SARS-CoV-2 variants (28). Briefly, we performed data QC, trim and quality filtering on raw FASTQ files using fastp. We discarded reads with an average Phred quality score below 30 and length shorter than 50 bp. We mapped the trimmed, human-filtered reads to the SARS-CoV-2 (Wuhan-Hu-1 reference, RefSeq ID: NC_045512.2) genome using bowtie (v.2.4.4) (29). We soft-clipped the ARTIC primers using iVar (v.1.3.1) (30) using default settings. Subsequently, we performed variant calling using iVar and generated consensus sequences using BCFtools (v.1.14) (31). We retained variants with a minimum base quality of 20, allele frequency of 0.25 and coverage of 10 reads. We annotated variants with SnpEff (v.5.0e) (32) and SnpSift (v.4.3) (33). All filtered variants were then compared against a manually curated list of known and epidemiologically important SARS-CoV-2 variants (including VOCs and VOIs) built using the 'outbreakinfo' R package (34).

We used Pangolin COVID-19 Lineage Assignment tool (v.4.0) to detect potential SARS-CoV-2 lineages from consensus sequences (35). We obtained a summary of raw reads, sample QC, alignment and variant calling metrics via MultiQC (v.1.11) (36). Unlike a clinical sample, each wastewater sample can contain multiple different SARS-CoV-2 virus lineages shed by many infected individuals. We used Freyja (v1.3.8) (25) to delineate and ascertain the relative proportion of different SARS-CoV-2 lineages in each wastewater sample.

Results

Viral RNA concentrations

Viral RNA concentrations varied widely over the study period from non-detect to a maximum of over 22,000 copies/mL. The maximum concentrations collected from partner facilities also varied by over an order of magnitude. The highest concentrations were observed at

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

the height of the Omicron wave (mid to late January 2022), while the lowest concentrations were observed shortly thereafter in mid to late March 2022 (Appendix Figure 2).

SARS-CoV-2 lineage diversity in wastewater *vs* clinical samples

Of the 1,235 samples collected from 32 sewersheds between November 2021 and March 2022, 1,067 had sufficient RNA for sequencing. The DRAGEN COVIDSeq pipeline successfully generated FASTQ files for 1,034 of those samples. We obtained a total of 9.86 ± 6.08 million raw reads and retained an average of 8.75 ± 5.37 million reads after quality control and read trimming. We found that on average 62.57 ± 25.45% of total reads mapped to the SARS-CoV-2 reference genome, and covered an average 87.37 ± 18.94% of the SARS-CoV-2 genome at 1X across all wastewater samples (Appendix Figures 3 and 4). We detected a total of 70 different lineages (including sub-lineages and with frequency > 5) from 1,034 consensus sequences, following PANGO nomenclature. Analysis of diversity of SARS-CoV-2 lineages revealed Delta as the most frequently detected lineage in our dataset (frequency, n=1,621, 67.71%) during November, prior to the emergence of Omicron (Figure 3a). However, we first detected Omicron (B.1.1.529) in a wastewater sample collected on November 19th, 2021 from Hyrum City WWTP (HCWWTP21) at an abundance of 0.1% (Figure 3b). The first clinical case of Omicron in Utah was reported on November 24th, 2021 with a specimen collection date of November 18th, 2021, though it was travel-related and community transmission of Omicron in Utah was observed around mid-December (Figure 3c-d).

We observed a steady decrease in the proportion of the Delta (B.1.617.2) lineages and the onset of Omicron (B.1.1529) and its sub-lineage BA.1 (n=112, 6.79%) in wastewater by December. A simple timeline of the detection of Omicron lineages is shown in Figure 2. We found BA.1.9 (lineage abundance: 8.35%) from Snyderville Basin Silver Creek (SBWRDSC26)

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

on December 7th, 2021, followed by detection of BA.1.17 (lineage abundance: 4.60%) and BA.1.1.11 (lineage abundance: 27.93%) from Central Valley WRF (CVWRF13) on December 10th (Figure 4). By December 14th, Omicron spread to two other communities, Snyderville Basin East Canyon (SBWRDEC27; lineages BA.1.1: 15.8%, BA.1.9: 9.11%) and Brigham City (BCSD20; lineage BA.1.9: 9.82%), along with increased prevalence in Snyderville Basin Silver Creek (SBWRDSC26; lineages B.1.1.529: 9.41%, BA.1.9: 52.75% and BA.1.1.16: 14.58%). Over the next two weeks, several BA.1 sub-lineages of Omicron had spread to nearly all communities (28/32), including the then common BA.1.15 (Figure 4). By January 4th, prevalence of Omicron in wastewater samples had reached 57.58%, and it completely displaced Delta by February 7th, 2022 (Figure 3a). Between January and March, Omicron accounted for 95% of total lineages in circulation (Figure 3b), with BA.1 being the most common (n=3,037, 82.93%), followed by BA.2 (n=156, 4.26%) and BA.3 (n=7, 0.19%). Overall, we detected a comparatively rapid shift from the earlier predominant Delta lineage to Omicron in our wastewater data.

Our wastewater genomic surveillance also identified changes in abundance of other Omicron sub-lineages over time. We first detected the Omicron sub-lineage BA.2 in our wastewater sequencing data from Ash Creek SSD (ACSSD32) on January 21st, 2022 (lineage abundances: BA.2.3, 11.28% and BA.2, 1.79%; Figure 4). By March 29th, the proportion of BA.2 reached about 45% including its sub-lineages BA.2.1, BA.2.2, BA.2.3, BA.2.4, BA.2.5, BA.2.6, BA.2.7 and BA.2.8. On February 22nd, 2021, we first detected Omicron sub-lineage BA.3 (lineage abundance: 6.69%) in Orem WRF (OWRF7). Interestingly, Omicron BA.3 persisted at low abundance in wastewater samples through March, but the BA.3 sub-lineage has not yet been identified in clinical surveillance. The patterns observed in wastewater sequencing

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

data were reflected in clinical sequencing (Figure 3c) and resulted in a huge surge in Omicron cases in December-February (Appendix Figure 1b). The increased transmission and spike in clinical cases due to Omicron was also evident in wastewater RNA viral loads (Appendix Figure 1a) and later also observed in our analysis of the proportion of mutations associated with the Omicron lineage.

Mutational spectra of SARS-CoV-2 in wastewater

We found S:D614G, ORF1ab:T3255I, S:T478K, S:N969K and E:T9I as the most common mutations among all wastewater samples (Appendix Figure 5). All five mutations were associated with known VOCs. Notably, all wastewater samples prior to the onset of Omicron (during November 2021) had signature mutations linked to the predominant Delta variant (e.g., S:P681R, S:D950N, T19R), followed by an increase in frequency of mutations shared between Delta and Omicron lineages (e.g., S:L452R, S:T478K; Appendix Figure 6).

An in-depth analysis of mutations from wastewater samples revealed an increase in frequency of several Omicron specific mutations (e.g., ORF1ab:T3255I, S:T95I, S:A67V, S:N856K, S:G142D; allele frequency >0.25) in late December 2021 to January 2022 (Figure 5). Genomic sequencing for wastewater samples collected from Roosevelt City (ROOSCD30) on November 2nd, 2021 confirmed the presence of five signature Omicron mutations (E:T9I, S:G339D, S:371L, S:373P, S:375P). We also identified two Spike protein mutations, A67V and T95I (allele frequency 1.00), specific to the Omicron lineage BA.1 from Provo City (PCWRF8) on November 30th, 2021. Later, we identified six signature mutations of Omicron (ORF1ab:T3255I, S:T95I, S:G142D, S:A67V, S:N856K, S:D796Y) in Snyderville Basin Silver Creek (SBWRDSC26) sampled on December 7th, 2021, supporting the assignment of Omicron BA.1 in our lineage analysis (Figure 4). We also identified 25 signature mutations of Omicron

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

from Central Valley WRF (CVWRF13) sampled on December 10th, 2021, aligning with its assignment as Omicron BA.1 (Figure 5).

Discussion

As the SARS-CoV-2 virus continues to spread and evolve rapidly to form new variants with increased virulence or immune escape potential, it is important to monitor and characterize VOI/VOCs. This is especially important now when pandemic fatigue has become widespread, clinical testing is reduced and more at-home tests are utilized. These factors together contribute towards delays in detecting new emerging lineages when VOCs may be already circulating in communities. In such scenarios, wastewater genomic surveillance could be essential to capture community transmission of COVID in a rapid, cost-effective and unbiased manner. It can provide a more comprehensive picture of SARS-CoV-2 genomic diversity at the community level than individual level data obtained by clinical sequencing. This will allow timely detection of newly emerging lineages and support public health efforts in mitigating COVID-19 spread (37).

Our wastewater sequencing data provides a first detailed view of the diversity of SARS-CoV-2 lineages circulating between November 2021 and March 2022 in Utah, encompassing the displacement of Delta by Omicron lineages. The trends in SARS-CoV-2 RNA levels across sewersheds accurately captured the surge in clinical cases among Utah's population (Appendix Figure 1). Notably, we found that the prevalence and diversity of SARS-CoV-2 lineages derived from wastewater samples closely aligned with the lineage proportions derived from clinical genomic surveillance. Sequencing of both wastewater samples and clinical samples between December and January revealed a notable increase in the prevalence of Omicron BA.1 lineages and corresponding decrease in the prevalence of Delta lineages (Figure 3). Omicron BA.1 was

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

the predominant lineage from January through early March, aligned with its global surge in cases.

One of the main advantages of wastewater surveillance is that it can be used for early detection of emerging lineages and help implement public health interventions to limit variant spread (16,17). In our study, comparing the collection date of Omicron-positive samples from wastewater with the collection dates of samples from clinical genomic surveillance, we detected Omicron lineages in wastewater genomic surveillance up to seven days before corresponding detections through genomic clinical surveillance. The first few clinical cases detected in Utah were travel-related to a region with high community transmission of Omicron. The lead time between the detection of a new lineage in wastewater compared to clinical surveillance is still difficult to estimate in advance and has been shown to vary from 5-14 days (12,16,17,38). The lead time can potentially be influenced by environmental factors– temperature, seasons (39), biological factors such as the shedding rates associated with a lineage (40), severity of infection and whether an individual is symptomatic or asymptomatic (38) when contributing to a particular catchment.

In our study, analysis of wastewater samples by Freyja, which provides relative abundance of multiple lineages in a sample, we first detected Omicron (B.1.1.529) from a Hyrum City (HCWWTP) sample collected on November 19th, 2021. However, based on Pangolin lineage assignment on consensus sequence, we first detected the Omicron lineage (BA.1.1) on December 13th, 2021 in Snyderville Basin Silver Creek. This highlights the uncertainty in lineage assignment based on consensus sequences derived from wastewater samples. Pangolin lineage assignment tools are perhaps better at identifying the predominant viral lineage in a

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

community but unlikely to represent the true lineage diversity as is expected in a wastewater sample, an important consideration for future studies.

Throughout the duration of our study encompassing the peak transmission of Omicron, wastewater sequencing identified three sub-lineages of Omicron: BA.1, BA.2 and BA.3. BA.1 was detected more frequently during our study compared to BA.2 across all sampled communities. By the end of March, as Omicron spread progressed, we observed an increased prevalence of BA.2 (45% prevalence on March 29th), similar to its prevalence in other states in the US. While BA.1 was still dominant across communities, we first detected BA.3 from Orem WRF in February. Surprisingly, BA.3 has not been detected in clinical surveillance in Utah. This could possibly be due to its low prevalence as was observed in our wastewater data (Figure 3b) and has been observed throughout the US (<0.5% prevalence, last checked Sep 19, 2022 on outbreak.info; (41)), suggesting its cryptic circulation in Utah's wastewater communities. These findings are in agreement with previous studies that highlight the utility of wastewater surveillance in detection of cryptic lineages – novel SARS-CoV-2 lineages that are rarely observed in humans (4,25).

Interestingly, the shift from Delta to Omicron was also quite clear from SARS-CoV-2 mutations in wastewater. We found predominance of Delta mutations in November-early December, followed by an increase in the number of mutations common between Delta and Omicron lineages and then an increase in Omicron defining mutations (Figure 4 and Appendix Figure 6). Thus, our ongoing wastewater genomic surveillance allowed us to detect Omicron lineage even when based on its signature mutations (e.g., S:A67V, S:N856K, S:N969K. E:T9I). Our study suggests that early detection of a lineage is feasible based on signature mutations, however, we caution against the interpretation of lineages based on individual mutations,

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

especially during the early stages of arrival of a new lineage. This is because frequencies of signature mutations associated with a new lineage are generally low, making it difficult to filter out the noise from the actual signal in the sequencing data.

Additionally, confident assignment of lineages for a wastewater sample requires availability of a comprehensive and well-curated lineage database. Novel lineages that have recently diverged from its ancestral lineage and not yet been curated in the SARS-CoV-2 lineage databases (e.g. Pangolin) will be challenging to identify via wastewater sequencing. However, analysis of key mutations from a clinically important region of the SARS-CoV-2 genome (e.g. Spike gene) that can influence transmissibility, immune or vaccine escape could help us identify a potential new lineage in wastewater. Recent studies have reported the presence of several rare, cryptic lineages in wastewater – SARS-CoV-2 lineages that did not match to any known lineages and had not been reported in humans (4,42). These novel lineages contained several rare mutations that had not been identified in routine clinical surveillance and shared many mutations with the circulating VOCs. Routine monitoring of such mutations of concern from wastewater communities could reveal new SARS-CoV-2 lineages before they become epidemiologically relevant and spread to the wider human population, thus guiding public health decision-making.

Taken together, our study illustrates that genomic sequencing of wastewater samples can be used to detect, monitor and provide insights into the diversity of SARS-CoV-2 variants. Of note is the cost-effectiveness of wastewater sequencing as a pooled community sample can be analyzed to monitor disease spread. WBE can be used to track the prevalence of SARS-CoV-2 variants in near real-time, often earlier than information obtained by clinical sample sequencing. Early detection of emerging lineages can provide the time needed to allocate more testing resources to communities showing early detection of an emerging variant and implement public

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

health interventions to limit variant spread. As we move beyond Omicron and new variants emerge, wastewater-based surveillance will be a valuable tool in the timely identification of novel lineages, complementing clinical sequencing and aiding in public health decision-making. While wastewater surveillance has proven to be an essential tool for SARS-CoV-2 genomic surveillance, it would be worthwhile expanding its application to detect other emerging infections and aid surveillance of future emerging infectious diseases.

Acknowledgments

We thank our partner wastewater treatment facilities across the State of Utah who collected and sent us the wastewater samples. This work was supported by the Centers for Disease Control and Prevention's (CDC) Epidemiology and Laboratory Capacity for Prevention and Control of Emerging Infectious Diseases (ELC) Enhanced Detection Cooperative Agreement.

Data Availability

The sequence data generated in the present study has been submitted to NCBI Sequence Read Archive and available under BioProject PRJNA812604.

References

1. Cameroni E, Bowen JE, Rosen LE, Saliba C, Zepeda SK, Culap K, et al. Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. Nature [Internet]. 2022 Feb 24 [cited 2022 Sep 20];602(7898):664–70. https://www.nature.com/articles/s41586-021-04386-2
2. Plante JA, Mitchell BM, Plante KS, Debbink K, Weaver SC, Menachery VD. The variant gambit: COVID-19's next move. Cell Host Microbe [Internet]. 2021 Apr 14;29(4):508–15. https://www.sciencedirect.com/science/article/pii/S1931312821000998
3. Kumar M, Patel AK, Shah AV, Raval J, Rajpara N, Joshi M, et al. First proof of the capability of wastewater surveillance for COVID-19 in India through detection of genetic material of

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

SARS-CoV-2. Sci Total Environ [Internet]. 2020 Dec 1;746:141326. https://www.sciencedirect.com/science/article/pii/S0048969720348555

4. Smyth DS, Trujillo M, Gregory DA, Cheung K, Gao A, Graham M, et al. Tracking cryptic SARS-CoV-2 lineages detected in NYC wastewater. Nat Commun [Internet]. 2022 Dec [cited 2022 Sep 1];13(1):635. https://www.nature.com/articles/s41467-022-28246-3
5. Ahmed W, Angel N, Edson J, Bibby K, Bivins A, O'Brien JW, et al. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. Sci Total Environ [Internet]. 2020 Aug 1;728:138764. https://www.sciencedirect.com/science/article/pii/S0048969720322816
6. Haramoto E, Malla B, Thakali O, Kitajima M. First environmental surveillance for the presence of SARS-CoV-2 RNA in wastewater and river water in Japan. Sci Total Environ [Internet]. 2020 Oct 1;737:140405. https://www.sciencedirect.com/science/article/pii/S0048969720339279
7. Rothman JA, Loveless TB, Kapcia J, Adams ED, Steele JA, Zimmer-Faust AG, et al. RNA Viromics of Southern California Wastewater and Detection of SARS-CoV-2 Single-Nucleotide Variants. Nojiri H, editor. Appl Environ Microbiol [Internet]. 2021 Nov 10 [cited 2022 Nov 22];87(23):e01448-21. https://journals.asm.org/doi/10.1128/AEM.01448-21
8. Asghar H, Diop OM, Weldegebriel G, Malik F, Shetty S, El Bassioni L, et al. Environmental Surveillance for Polioviruses in the Global Polio Eradication Initiative. J Infect Dis [Internet]. 2014 Nov 1 [cited 2022 Aug 1];210(suppl_1):S294–303. https://doi.org/10.1093/infdis/jiu384
9. The Global Sewage Surveillance project consortium, Hendriksen RS, Munk P, Njage P, van Bunnik B, McNally L, et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nat Commun [Internet]. 2019 Dec [cited 2022 Aug

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

1];10(1):1124. http://www.nature.com/articles/s41467-019-08853-3

10. González Mariño I, Baz Lomba JA, Alygizakis NA, Andrés□Costa MJ, Bade R, Bannwarth A, et al. Spatio□temporal assessment of illicit drug use at large scale: evidence from 7 years of international wastewater monitoring. Addiction [Internet]. 2020 Jan [cited 2022 Aug 1];115(1):109–20. https://onlinelibrary.wiley.com/doi/10.1111/add.14767

11. Peccia J, Zulli A, Brackney DE, Grubaugh ND, Kaplan EH, Casanovas-Massana A, et al. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. Nat Biotechnol [Internet]. 2020 Oct [cited 2022 Sep 1];38(10):1164–7. https://www.nature.com/articles/s41587-020-0684-z

12. Medema G, Heijnen L, Elsinga G, Italiaander R, Brouwer A. Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. Environ Sci Technol Lett [Internet]. 2020 Jul 14 [cited 2022 Sep 1];7(7):511–6. https://pubs.acs.org/doi/10.1021/acs.estlett.0c00357

13. Acosta N, Bautista MA, Waddell BJ, McCalder J, Beaudet AB, Man L, et al. Longitudinal SARS-CoV-2 RNA wastewater monitoring across a range of scales correlates with total and regional COVID-19 burden in a well-defined urban population. Water Res [Internet]. 2022 Jul [cited 2022 Sep 1];220:118611. https://linkinghub.elsevier.com/retrieve/pii/S0043135422005644

14. Weidhaas J, Aanderud ZT, Roper DK, VanDerslice J, Gaddis EB, Ostermiller J, et al. Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. Sci Total Environ [Internet]. 2021 Jun [cited 2022 Sep 1];775:145790. https://linkinghub.elsevier.com/retrieve/pii/S0048969721008573

15. Shrestha S, Yoshinaga E, Chapagain SK, Mohan G, Gasparatos A, Fukushi K. Wastewater-

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

Based Epidemiology for Cost-Effective Mass Surveillance of COVID-19 in Low- and Middle-Income Countries: Challenges and Opportunities. Water [Internet]. 2021 Oct 15 [cited 2022 Sep 9];13(20):2897. https://www.mdpi.com/2073-4441/13/20/2897

16. Ahmed W, Tscharke B, Bertsch PM, Bibby K, Bivins A, Choi P, et al. SARS-CoV-2 RNA monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: A temporal case study. Sci Total Environ [Internet]. 2021 Mar [cited 2022 Aug 1];761:144216. https://linkinghub.elsevier.com/retrieve/pii/S0048969720377470

17. Karthikeyan S, Ronquillo N, Belda-Ferre P, Alvarado D, Javidi T, Longhurst CA, et al. High-Throughput Wastewater SARS-CoV-2 Detection Enables Forecasting of Community Infection Dynamics in San Diego County. Cristea IM, editor. mSystems [Internet]. 2021 Apr 27 [cited 2021 Aug 5];6(2). https://journals.asm.org/doi/10.1128/mSystems.00045-21

18. Wurtzer S, Waldman P, Ferrier-Rembert A, Frenois-Veyrat G, Mouchel JM, Boni M, et al. Several forms of SARS-CoV-2 RNA can be detected in wastewaters: Implication for wastewater-based epidemiology and risk assessment. Water Res [Internet]. 2021 Jun [cited 2022 Sep 7];198:117183. https://linkinghub.elsevier.com/retrieve/pii/S004313542100381X

19. Wolfe M, Hughes B, Duong D, Chan-Herur V, Wigginton KR, White BJ, et al. Detection of SARS-CoV-2 Variants Mu, Beta, Gamma, Lambda, Delta, Alpha, and Omicron in Wastewater Settled Solids Using Mutation-Specific Assays Is Associated with Regional Detection of Variants in Clinical Samples. Buan NR, editor. Appl Environ Microbiol [Internet]. 2022 Apr 26 [cited 2022 Sep 9];88(8):e00045-22. https://journals.asm.org/doi/10.1128/aem.00045-22

20. Heijnen L, Elsinga G, Graaf M de, Molenkamp R, Koopmans MPG, Medema G. Droplet digital RT-PCR to detect SARS-CoV-2 signature mutations of variants of concern in

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

wastewater. Sci Total Environ [Internet]. 2021;799:149456. https://www.sciencedirect.com/science/article/pii/S0048969721045307

21. Amman F, Markt R, Endler L, Hupfauf S, Agerer B, Schedl A, et al. Viral variant-resolved wastewater surveillance of SARS-CoV-2 at national scale. Nat Biotechnol [Internet]. 2022 Jul 18 [cited 2022 Sep 1]; https://www.nature.com/articles/s41587-022-01387-y

22. Resende PC, Motta FC, Roy S, Appolinario L, Fabri A, Xavier J, et al. SARS-CoV-2 genomes recovered by long amplicon tiling multiplex approach using nanopore sequencing and applicable to other sequencing platforms [Internet]. Molecular Biology; 2020 May [cited 2021 Aug 5]. http://biorxiv.org/lookup/doi/10.1101/2020.04.30.069039

23. Crits-Christoph A, Kantor RS, Olm MR, Whitney ON, Al-Shayeb B, Lou YC, et al. Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants. 2021;12(1):9.

24. Rehn A, Braun P, Knüpfer M, Wölfel R, Antwerpen MH, Walter MC. Catching SARS-CoV-2 by Sequence Hybridization: a Comparative Analysis. mSystems [Internet]. [cited 2021 Aug 10];0(0):e00392-21. https://journals.asm.org/doi/10.1128/mSystems.00392-21

25. Karthikeyan S, Levy JI, De Hoff P, Humphrey G, Birmingham A, Jepsen K, et al. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. Nature [Internet]. 2022 Sep 1 [cited 2022 Sep 1];609(7925):101–8. https://www.nature.com/articles/s41586-022-05049-6

26. Baaijens JA, Zulli A, Ott IM, Nika I, van der Lugt MJ, Petrone ME, et al. Lineage abundance estimation for SARS-CoV-2 in wastewater using transcriptome quantification techniques. Genome Biol [Internet]. 2022 Nov 8 [cited 2022 Nov 17];23(1):236. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-022-02805-9

27. Centers for Disease Control and Prevention. CDC 2019-novel coronavirus (2019-nCoV) real-

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

time RT-PCR diagnostic panel. 2021;

28. Patel H, Varona S, Monzon S, Espinosa-Carrasco; J, L Heuer M, Underwood A, et al. nf-core/viralrecon: nf-core/viralrecon v2.4.1 [Internet]. 2022. (Zenodo). https://zenodo.org/record/6320980#.Y3a7sezMLJx
29. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol [Internet]. 2009 [cited 2022 Sep 16];10(3):R25. http://genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-3-r25
30. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol [Internet]. 2019 Dec [cited 2022 Sep 16];20(1):8. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-018-1618-7
31. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics [Internet]. 2011 Nov 1 [cited 2022 Oct 7];27(21):2987–93. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btr509
32. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w ^{1118}□; iso-2; iso-3. Fly (Austin) [Internet]. 2012 Apr [cited 2022 Oct 7];6(2):80–92. http://www.tandfonline.com/doi/abs/10.4161/fly.19695
33. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, et al. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet [Internet]. 2012 [cited 2022 Oct 7];3. http://journal.frontiersin.org/article/10.3389/fgene.2012.00035/abstract

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

34. Alkuzweny M, Hughes L, Gangavarapu K. outbreakinfo: outbreak.info R Client [Internet]. 2022. https://outbreak-info.github.io/R-outbreak-info/

35. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool. Virus Evol [Internet]. 2021 Jul 5 [cited 2022 Oct 7];veab064. https://academic.oup.com/ve/advance-article/doi/10.1093/ve/veab064/6315289

36. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics [Internet]. 2016 Oct 1 [cited 2022 Oct 7];32(19):3047–8. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw354

37. Kirby AE, Walters MS, Jennings WC, Fugitt R, LaCross N, Mattioli M, et al. Using Wastewater Surveillance Data to Support the COVID-19 Response — United States, 2020–2021. MMWR Morb Mortal Wkly Rep [Internet]. 2021 Sep 10 [cited 2022 Sep 1];70(36):1242–4. http://www.cdc.gov/mmwr/volumes/70/wr/mm7036a2.htm?s_cid=mm7036a2_w

38. D'Aoust PM, Graber TE, Mercier E, Montpetit D, Alexandrov I, Neault N, et al. Catching a resurgence: Increase in SARS-CoV-2 viral RNA identified in wastewater 48 h before COVID-19 clinical tests and 96 h before hospitalizations. Sci Total Environ [Internet]. 2021 May [cited 2022 Sep 15];770:145319. https://linkinghub.elsevier.com/retrieve/pii/S0048969721003867

39. Kumar M, Jiang G, Kumar Thakur A, Chatterjee S, Bhattacharya T, Mohapatra S, et al. Lead time of early warning by wastewater surveillance for COVID-19: Geographical variations and impacting factors. Chem Eng J [Internet]. 2022 Aug [cited 2022 Sep 15];441:135936.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

https://linkinghub.elsevier.com/retrieve/pii/S1385894722014346

40. Singanayagam A, Hakki S, Dunning J, Madon KJ, Crone MA, Koycheva A, et al. Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study. Lancet Infect Dis [Internet]. 2022 Feb [cited 2022 Sep 15];22(2):183–95. https://linkinghub.elsevier.com/retrieve/pii/S1473309921006484

41. Hughes L, Gangavarapu K, Latif AA, Mullen J, Alkuzweny M, Hufbauer E, et al. Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations [Internet]. In Review; 2022 Jun [cited 2022 Oct 4]. https://www.researchsquare.com/article/rs-1723829/v1

42. Gregory DA, Trujillo M, Rushford C, Flury A, Kannoly S, San KM, et al. Genetic Diversity and Evolutionary Convergence of Cryptic SARS-CoV-2 Lineages Detected Via Wastewater Sequencing [Internet]. 2022 Jun [cited 2022 Oct 3]. http://medrxiv.org/lookup/doi/10.1101/2022.06.03.22275961

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

Figures

Figure 1: Geographic map showing the 32 sampled wastewater treatment sites in Utah.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

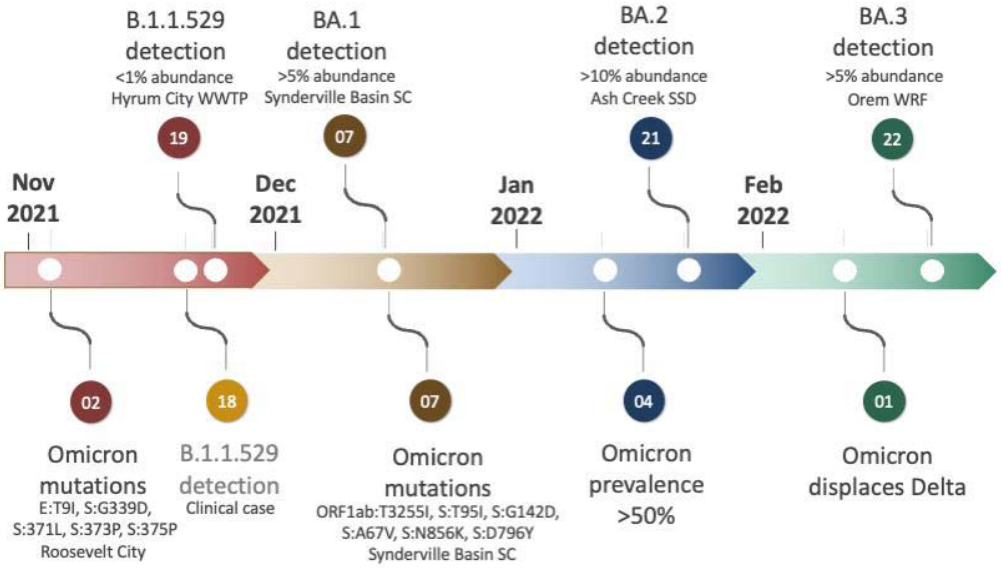


Figure 2 - Timeline of the important events in the detection of lineages and key mutations of Omicron from wastewater sequencing.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

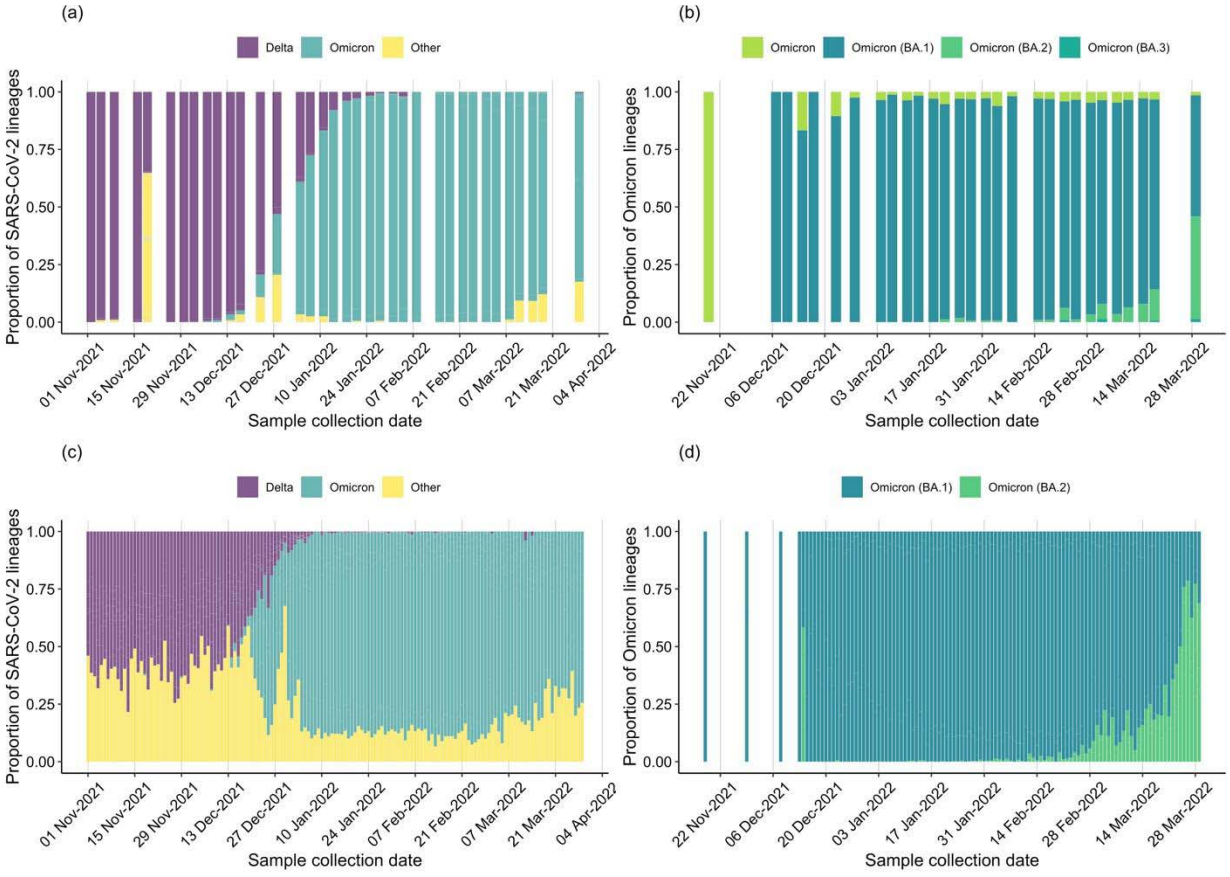


Figure 3 - Proportion of all Sars-CoV-2 lineages in a) wastewater samples, c) clinical samples and proportion of Omicron lineages in b) wastewater samples, d) clinical samples from November 2021 to March 2022.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

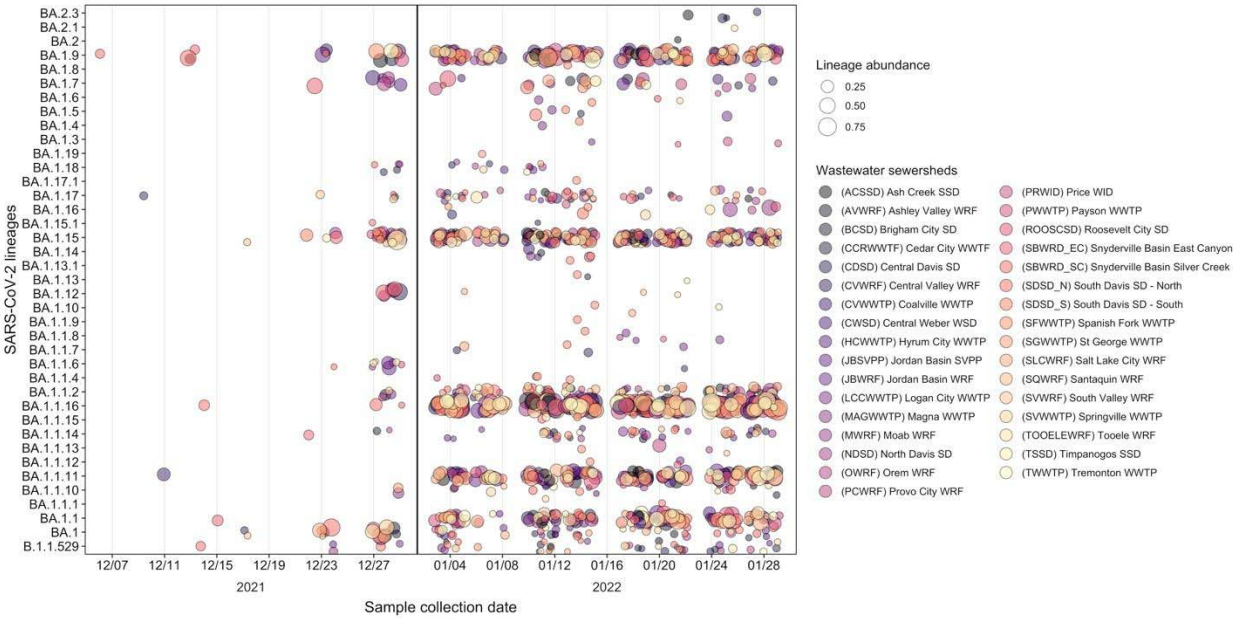


Figure 4 - Abundance of Omicron lineages and its sub-lineages in each sewershed between December 2021 and January 2022 to highlight the introduction and spread of Omicron in Utah. Circles are color coded by sewersheds and the size of each circle corresponds to the relative abundance of lineages

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

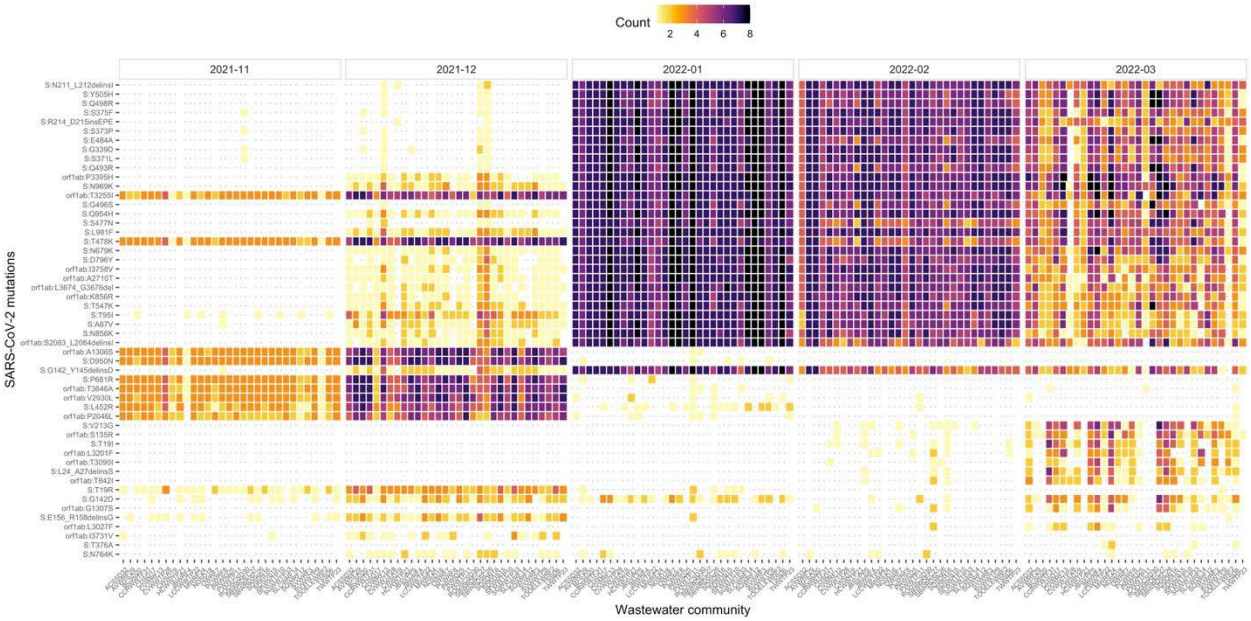
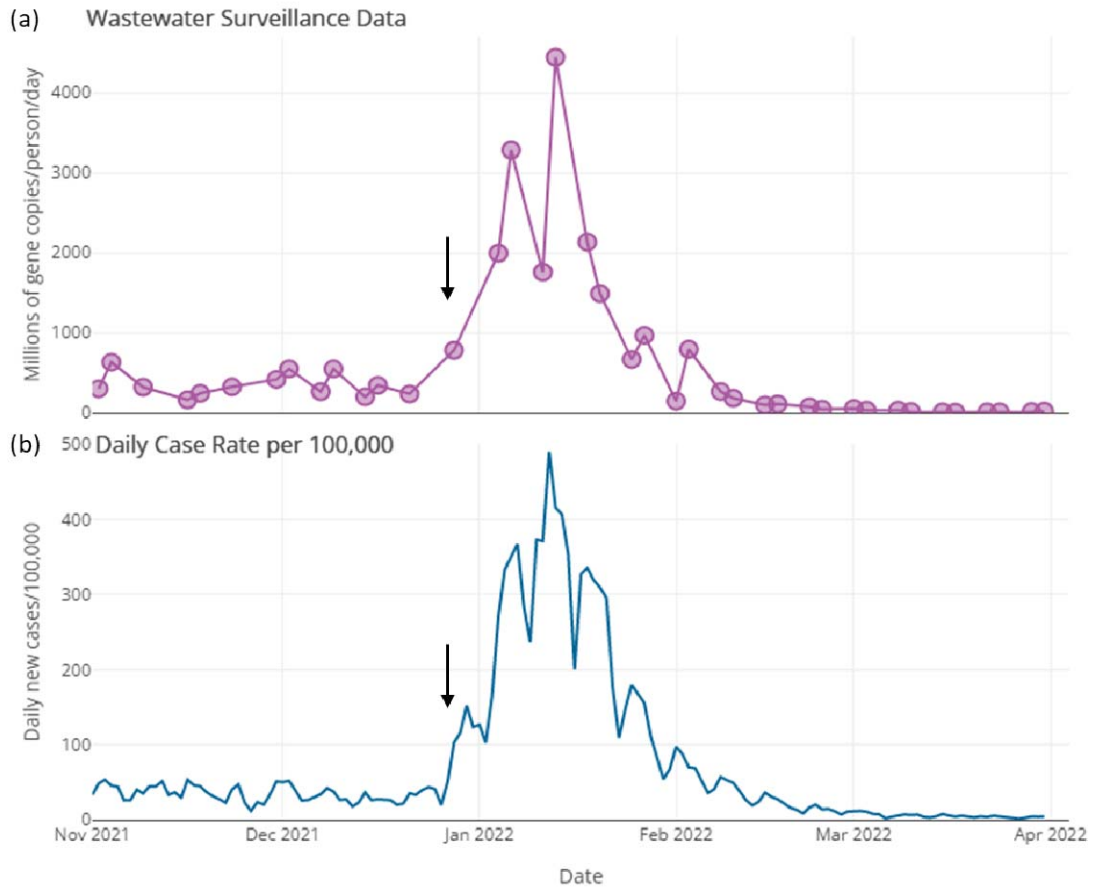


Figure 5- Heatmap showing the temporal changes in the frequency of SARS-CoV-2 amino acid changes in the Spike and ORF1ab genes, detected in each sewershed from November 2021 to March 2022, encompassing the onset of Omicron. Amino acid changes that were detected only once during the study period are not shown. Colors in the heatmap range from cool yellows/reds representing low frequency to warm blues/violets representing high frequency.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .

Supplementary Figures:

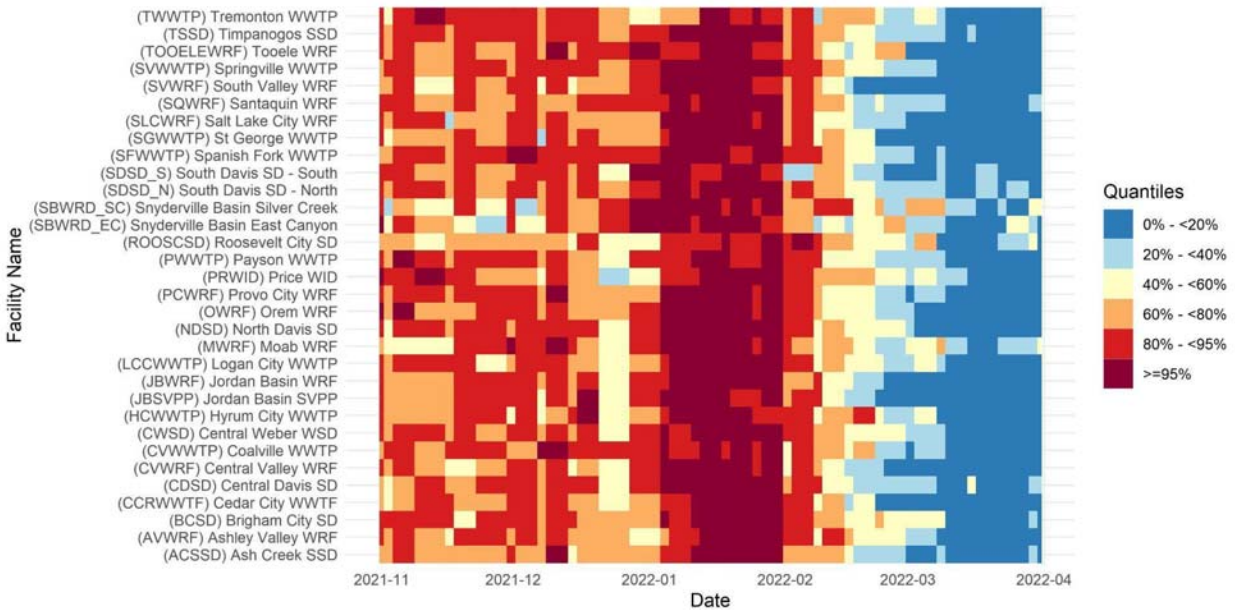


Appendix Figure 1: Example wastewater surveillance data and sewershed-associated case rates. The displayed data are from the (CVWRF) Central Valley WRF partner facility, and are generally representative of data from the time period included in the study. Vertical black arrows indicate the approximate start of the Omicron wave in late December 2021.

(a) Wastewater surveillance data. Data generated by UPHL in gene copies/mL of raw wastewater was normalized by estimated population and average flow during the sampling period (Appendix 1). (b) Daily sewershed associated case rates per 100,000. All COVID-19 cases in Utah are geocoded by sewershed.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

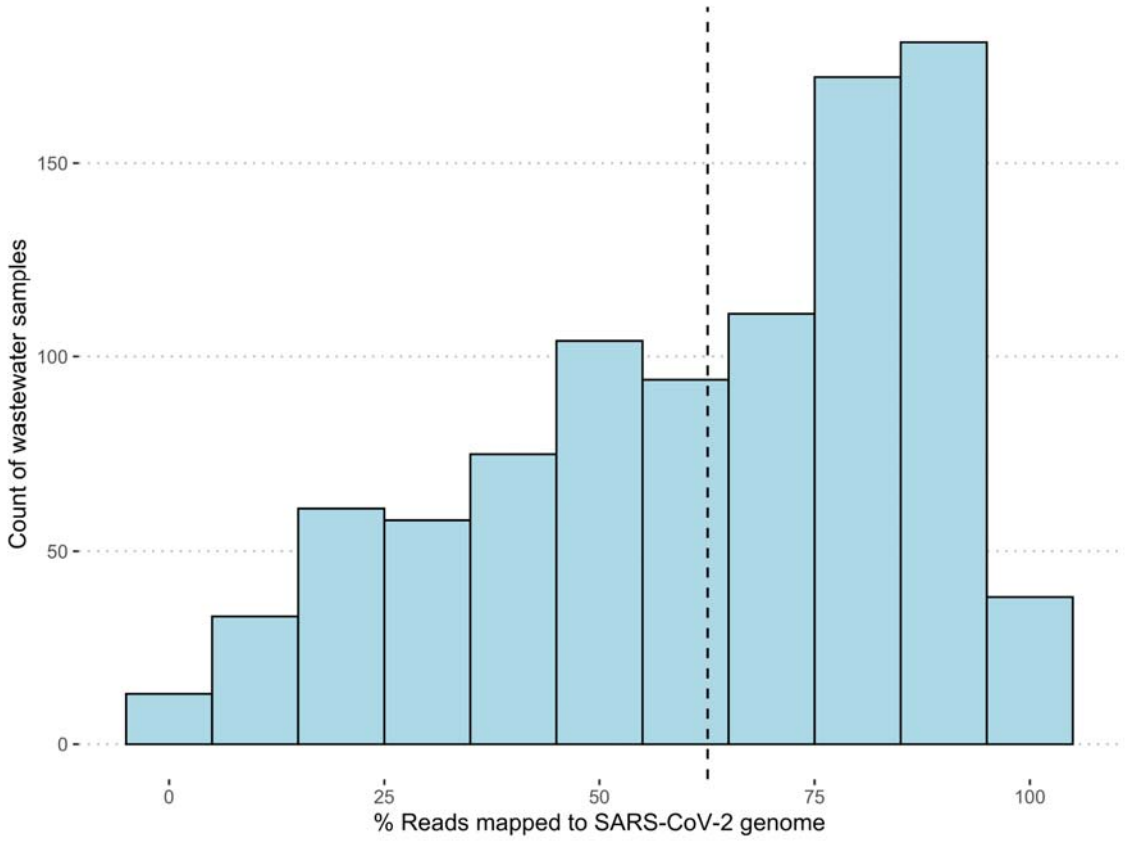
It is made available under a CC-BY-ND 4.0 International license .



Appendix Figure 2: Heatmap of wastewater concentration quantiles. SARS-CoV-2 concentrations in wastewater (normalized by flow and population) were categorized into quantile bins. Bin thresholds are standard quintiles with the addition of a 95%+ bin to better emphasize the highest concentrations. The numeric thresholds were calculated on a site-specific basis on all data from July 1, 2020 to September 22, 2022, although only data relevant to the current manuscript is displayed.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

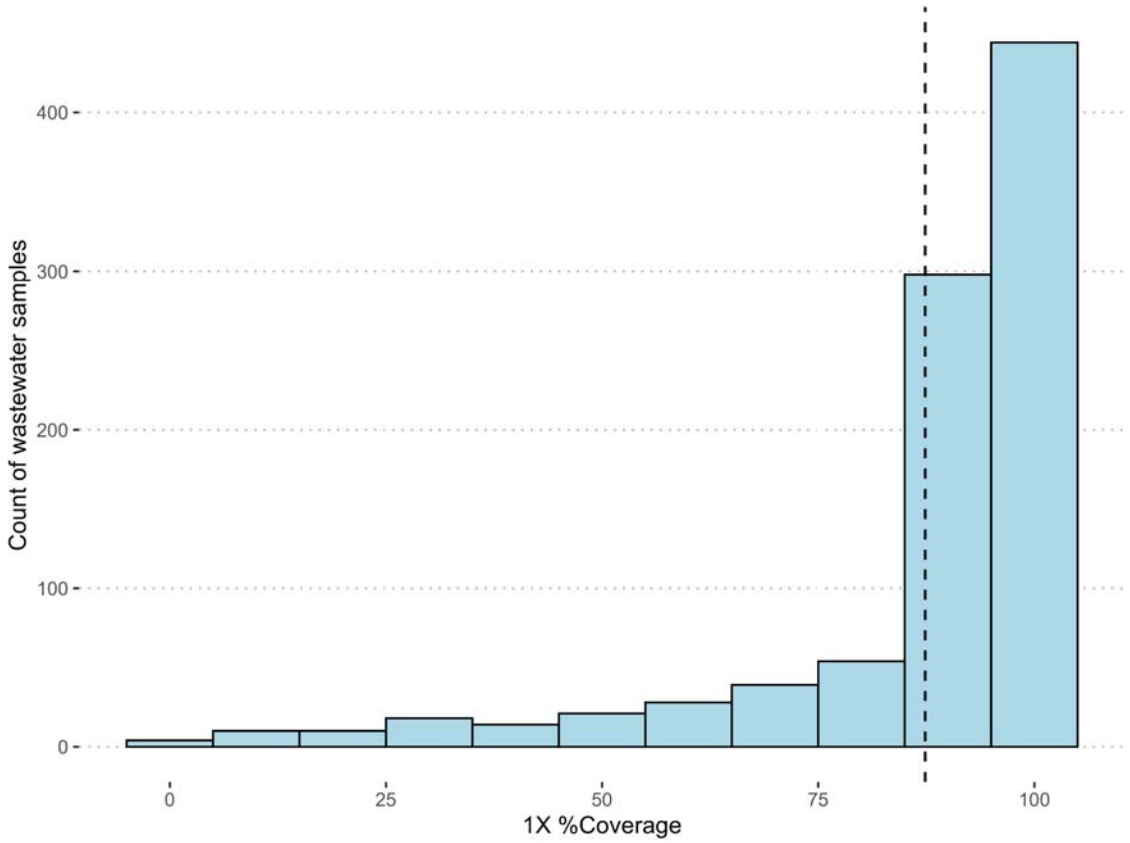
It is made available under a CC-BY-ND 4.0 International license .



Appendix Figure 3 - Histogram of the number of wastewater samples and their percent reads mapped to the SARS-CoV-2 genome. The black dashed line indicates the mean percentage of mapped reads.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

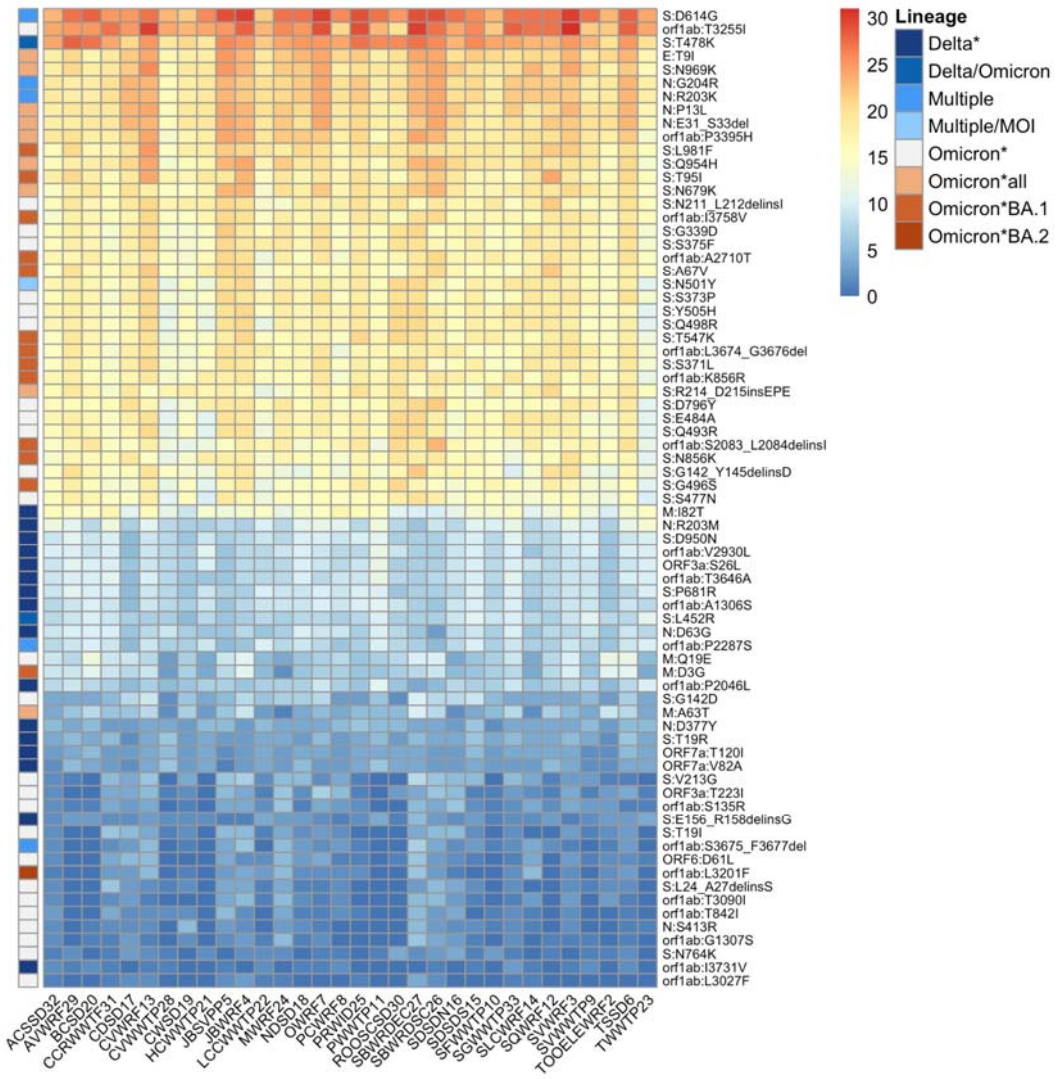
It is made available under a CC-BY-ND 4.0 International license .



Appendix Figure 4 - Histogram of the number of wastewater samples and their 1X coverage. The black dashed line indicates the mean 1X coverage.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

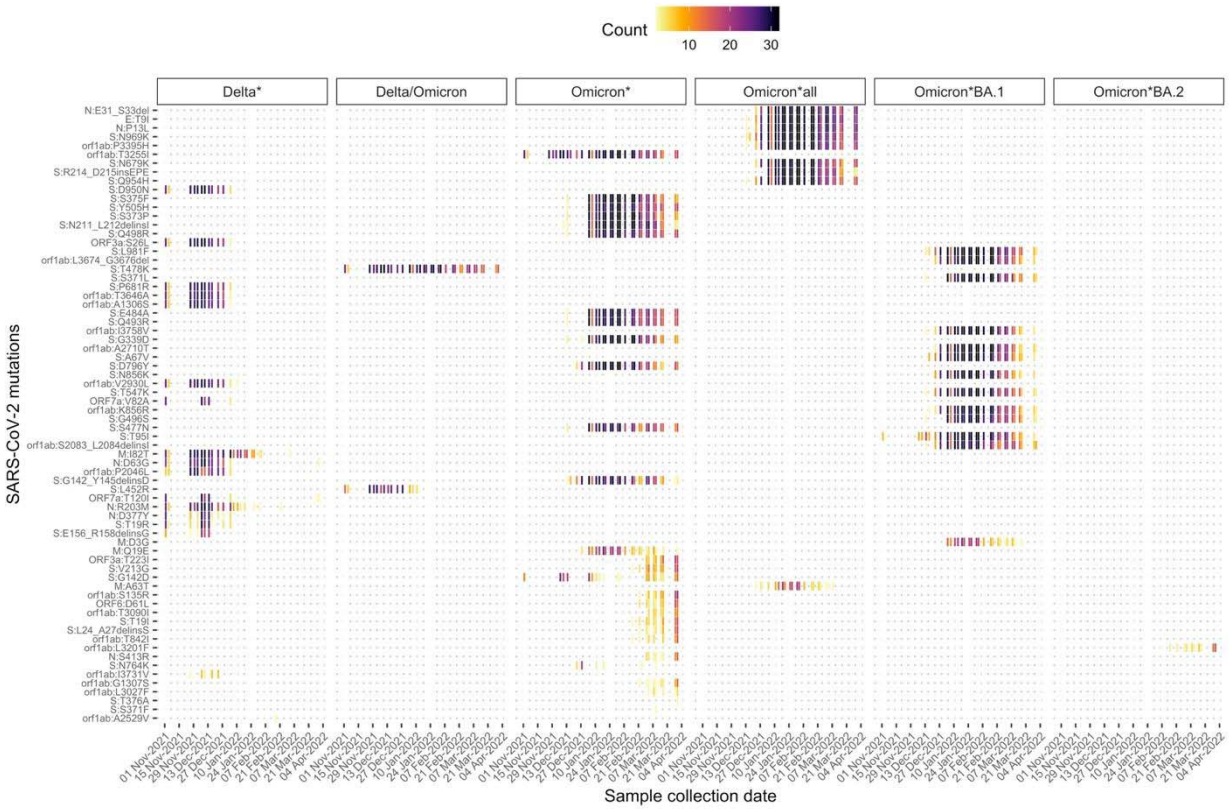
It is made available under a CC-BY-ND 4.0 International license .



Appendix Figure 5- Heatmap of the frequency of all amino acid changes during November 2021 to March 2022 for each sewershed. Colors in the matrix range from cool blues representing low frequency to warm orange representing high frequency. The amino acid changes specific to a SARS-CoV-2 lineage are color coded on the left bar of the heatmap – 'Multiple' refers to mutations shared between multiple SARS-CoV-2 lineages and 'Multiple/MOI' refers to 'mutations of interest' shared between multiple SARS-CoV-2 lineages. Data shown here has been truncated to amino acid changes that were detected 10 or more times during the study period.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.24.22282643; this version posted November 29, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-ND 4.0 International license .



Appendix Figure 6 - Heatmap showing the temporal changes in the frequency of SARS-CoV-2 amino acid changes from November 2021 to March 2022. Amino acid changes that were detected only once during the study period are not shown. Colors in the matrix range from light yellow/reds representing low frequency to dark blues/violet representing high frequency.