Check for updates

RNA Viromics of Southern California Wastewater and Detection of SARS-CoV-2 Single-Nucleotide Variants

Jason A. Rothman,[a] Theresa B. Loveless,[b] Joseph Kapcia III,[a] Eric D. Adams,[a] Joshua A. Steele,[c] Amity G. Zimmer-Faust,[c] Kylie Langlois,[c] David Wanless,[c] Madison Griffith,[c] Lucy Mao,[c] Jeffrey Chokry,[c] John F. Griffith,[c] Katrine L. Whiteson[a]

aDepartment of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, California, USA
bDepartment of Biomedical Engineering, University of California, Irvine, Irvine, California, USA
cSouthern California Coastal Water Research Project, Costa Mesa, California, USA

ABSTRACT Municipal wastewater provides an integrated sample of a diversity of human-associated microbes across a sewershed, including viruses. Wastewater-based epidemiology (WBE) is a promising strategy to detect pathogens and may serve as an early warning system for disease outbreaks. Notably, WBE has garnered substantial interest during the coronavirus disease 2019 (COVID-19) pandemic to track disease burden through analyses of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA. Throughout the COVID-19 outbreak, tracking SARS-CoV-2 in wastewater has been an important tool for understanding the spread of the virus. Unlike traditional sequencing of SARS-CoV-2 isolated from clinical samples, which adds testing burden to the health care system, in this study, metatranscriptomics was used to sequence virus directly from wastewater. Here, we present a study in which we explored RNA viral diversity through sequencing 94 wastewater influent samples across seven wastewater treatment plants (WTPs), collected from August 2020 to January 2021, representing approximately 16 million people in Southern California. Enriched viral libraries identified a wide diversity of RNA viruses that differed between WTPs and over time, with detected viruses including coronaviruses, influenza A, and noroviruses. Furthermore, single-nucleotide variants (SNVs) of SARS-CoV-2 were identified in wastewater, and we measured proportions of overall virus and SNVs across several months. We detected several SNVs that are markers for clinically important SARS-CoV-2 variants along with SNVs of unknown function, prevalence, or epidemiological consequence. Our study shows the potential of WBE to detect viruses in wastewater and to track the diversity and spread of viral variants in urban and suburban locations, which may aid public health efforts to monitor disease outbreaks.

IMPORTANCE Wastewater-based epidemiology (WBE) can detect pathogens across sewersheds, which represents the collective waste of human populations. As there is a wide diversity of RNA viruses in wastewater, monitoring the presence of these viruses is useful for public health, industry, and ecological studies. Specific to public health, WBE has proven valuable during the coronavirus disease 2019 (COVID-19) pandemic to track the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) without adding burden to health care systems. In this study, we used metatranscriptomics and reverse transcription-droplet digital PCR (RT-ddPCR) to assay RNA viruses across Southern California wastewater from August 2020 to January 2021, representing approximately 16 million people from Los Angeles, Orange, and San Diego counties. We found that SARS-CoV-2 quantification in wastewater correlates well with county-wide COVID-19 case data, and that we can detect SARS-CoV-2 single-nucleotide variants through sequencing. Likewise, wastewater treatment plants (WTPs) harbored different viromes, and we detected other human

Address correspondence to Jason A. Rothman, rothmanj@uci.edu, or Katrine L. Whiteson, katrine@uci.edu.

pathogens, such as noroviruses and adenoviruses, furthering our understanding of wastewater viral ecology.

Municipal wastewater represents a matrix containing a wide diversity of microbes and is representative of the collective waste of a human population across a catchment area. (1). The microbial content of wastewater can be useful in determining the levels of biological contamination of an area, including the presence of human and animal feces, antimicrobial resistance genes, pathogenic bacteria, and viruses (1–6). Regarding viruses specifically, wastewater often contains high titers of bacteriophages and plant-infecting viruses, along with generally smaller proportions of viruses that infect animals, including humans (7–10). Many studies have used metagenomics to characterize the viral content of wastewater, but these studies typically rely on extracted DNA, which is unable to capture the wide diversity of RNA-based viruses (11–13). As RNA viruses can be important pathogens of humans and agricultural organisms, using metatranscriptomic sequencing to study these diverse viruses in wastewater is relevant to public health and industry and may allow for a greater understanding of the ecological processes that occur in wastewater (2, 8, 14–16).

Wastewater-based epidemiology (WBE) is a useful method to detect the presence of human pathogens and may serve as an early warning system for disease outbreaks (17). For example, WBE has been used to track the prevalence of viruses such as norovirus, rotavirus, adenovirus, poliovirus, influenza, and severe acute respiratory syndrome (SARS) coronaviruses (18–23), with the added benefit that WBE does not rely on public health and clinical resources (17). Aside from basic detection, it has been shown that the viral load of wastewater often precedes clinical outbreaks and may offer a forecast of the severity of localized disease outbreaks (18, 24–26). Infectious diseases are often underreported in clinical settings, likely due to asymptomatic cases, avoidance of health care, and incorrect diagnoses, which prevents accurate disease surveillance and incidence reporting, potentially hampering public health responses (27, 28). While not all pathogens are excreted into wastewater at detectible levels, the presence of detectible viruses in wastewater and human fecal samples indicates that WBE may still provide useful information in monitoring disease (21, 29, 30).

The coronavirus disease 2019 (COVID-19) pandemic has placed an intense strain on health care systems worldwide and has resulted in over 4 million human deaths (31). Caused by the 2019 emergence of the enveloped positive-sense single-stranded RNA (+ssRNA) severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (32, 33), this virus has been reported in direct nasal swab patient samples, human feces, and wastewater samples (15, 23–25, 29, 34–36). While detection and quantification of SARS-CoV-2 are clearly important, the predominant method of detection, reverse transcription-quantitative PCR (RT-qPCR), does not allow for the characterization of viral variants, which is critical to monitor during the COVID-19 pandemic to track the evolution of SARS-CoV-2 (37, 38). In light of the limitations of RT-qPCR, thousands of patient-derived SARS-CoV-2 genomes have been sequenced, and important viral variants have been discovered (37, 39–41). However, the sequencing of patient samples relies mainly on clinical samples (42), which were difficult to obtain during the COVID-19 pandemic (43, 44). As wastewater represents a composite of human waste from the total catchment area, metagenomic sequencing of these samples may allow for the detection of viral variants and pathogen diversity across larger populations and regions without further burdening health care workers (11, 13, 45, 46).

Given the importance of monitoring the COVID-19 pandemic and exploring RNA viral diversity, in this study, metatranscriptomic sequencing and droplet digital PCR (ddPCR) were conducted in parallel to characterize the RNA viromes and viral load of SARS-CoV-2 of 94 influent samples from seven wastewater treatment plants (WTPs)
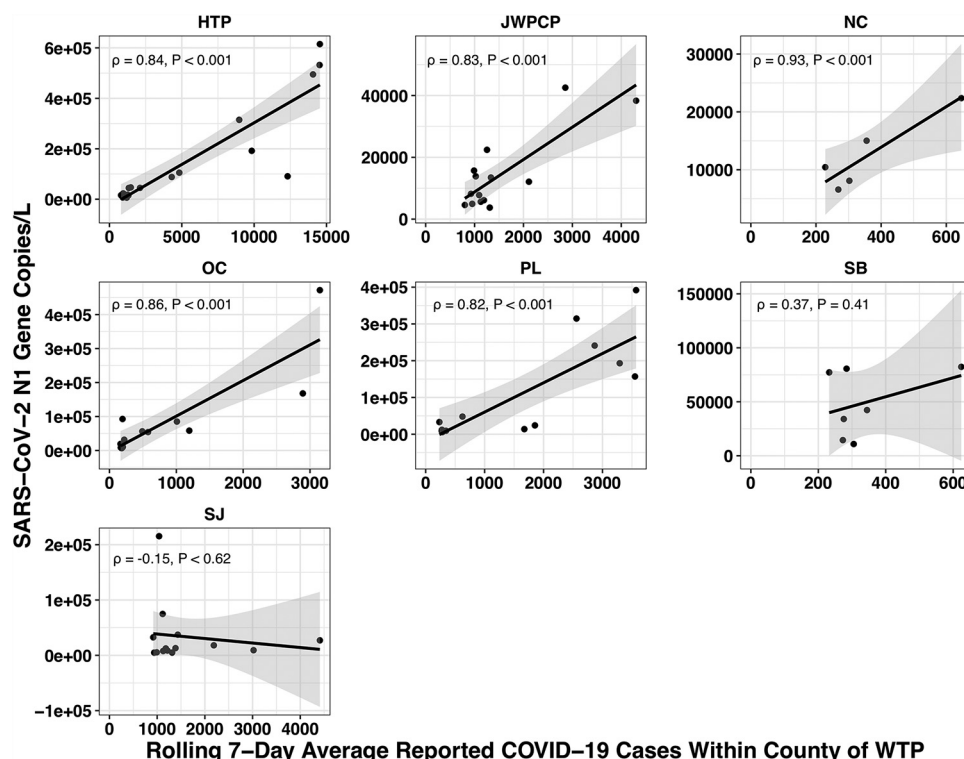
FIG 1 Pearson correlations of SARS-CoV-2 wastewater viral loads as measured by RT-ddPCR and the 7-day rolling average of COVID-19 case counts within county of WTP faceted by WTP. COVID-19 rolling averages and wastewater viral load were generally significantly correlated (HTP [$\rho = 0.84$, $P < 0.001$], JWPCP [$\rho = 0.83$, $P < 0.001$], NC [$\rho = 0.94$, $P = 0.03$], SB [$\rho = 0.37$, $P = 0.41$], OC [$\rho = 0.86$, $P < 0.001$], PL [$\rho = 0.82$, $P < 0.001$], and SJ [$\rho = -0.15$, $P = 0.62$]).

representing a total population of 16 million individuals across Southern California. Through our study, we investigated several lines of inquiry. First, what is the diversity of RNA viruses across Southern California wastewater, and does viral abundance change longitudinally? Second, can we detect human-infecting viruses, including coronaviruses, in wastewater, and how does SARS-CoV-2 quantification from wastewater samples compare to county-level COVID-19 case counts? Third, can we use metatranscriptomics to detect and track the emergence of SARS-CoV-2 strain variants in wastewater over time?

RESULTS

ddPCR quantification of SARS-CoV-2 and correlation with daily countywide COVID-19 cases. We quantified SARS-CoV-2 viral load in 85 influent wastewater samples from August 2020 to early January 2021 across seven WTPs in Los Angeles, Orange, and San Diego counties in California. Because of COVID-19 reporting at the county level, we correlated SARS-CoV-2 N1 gene copies per liter of influent wastewater with cases within counties only (i.e., the Hyperion [HTP] WTP was correlated with Los Angeles county COVID-19 cases) (Table S1 in the supplemental material). Generally, the viral load of each individual WTP correlated with the 7-day rolling average of reported countywide COVID-19 cases: HTP ($\rho = 0.84$, $P < 0.001$), Joint Water Pollution Control Plant (JWPCP) ($\rho = 0.83$, $P < 0.001$), North City (NC) ($\rho = 0.94$, $P = 0.03$), South Bay (SB) ($\rho = 0.37$, $P = 0.41$), Orange County (OC) ($\rho = 0.86$, $P < 0.001$), Point Loma (PL) ($\rho = 0.82$, $P < 0.001$), and San Jose Creek (SJ) (Los Angeles county, $\rho = -0.15$, $P = 0.62$) (Fig. 1).

Sequencing library characteristics. We obtained a total of 1,119,674,084 quality-filtered deduplicated paired reads across 180 libraries (90 unique samples), and, as we used two different library preparation strategies, we report the statistics separately. For

unenriched libraries, we obtained 794,773,512 reads (average: 84,55,037; range: 2,096,056 to 17,005,378), of which 40,512,123 (average: 430,980 [2.1%]; range: 3,560 [0.05%] to 1,675,525 [13.5%]) were viruses, and an average of 2.4% (range: 0.06% to 25.1%) were human reads. For enriched libraries, we obtained 324,900,572 reads (average: 3,777,914; range: 1,182,078 to 14,392,582), of which 11,660,490 (average: 135,587 [4.7%]; range: 2,580 [0.09%] to 484,944 [17.8%]) were viruses, and an average of 2.0% (range: 0.06% to 19.1%) were human reads. We were able to pair 172 libraries (86 unique samples) that had both unenriched and enriched samples successfully sequenced.

Viral composition of wastewater and detection of selected pathogens. As enrichment changed the viral composition of wastewater samples, we analyzed the unenriched and enriched sample data separately. Unenriched samples contained sequences from 2,495 viruses, and the top 10 most proportionally abundant viruses accounted for an average of 97.6% of the total abundance. The average relative abundances $\pm$ standard deviation of these "top 10" viruses were as follows: tomato brown rugose fruit virus (66.0 $\pm$ 7.4%), pepper mild mottle virus (PMMoV) (10.6 $\pm$ 3.7%), cucumber green mottle mosaic virus (10.4 $\pm$ 3.9%), tomato mosaic virus (4.8 $\pm$ 2.9%), tobacco mild green mosaic virus (2.1 $\pm$ 1.2%), tropical soda apple mosaic virus (1.5 $\pm$ 0.8%), tomato mottle mosaic virus (1.4 $\pm$ 1.8%), crAssphage (0.4 $\pm$ 0.8%), opuntia virus 2 (0.2 $\pm$ 0.4%), and melon necrotic spot virus (0.2 $\pm$ 0.2%) (Fig. 2).

We analyzed the data from Illumina respiratory virus (IRV)-enriched samples as above and found sequences from 2,215 viruses, with the top 10 most proportionally abundant viruses accounting for an average of 97.2% of total abundance. The average relative abundances $\pm$ standard deviation of these "top 10" viruses were as follows: tomato brown rugose fruit virus (60.2 $\pm$ 8.6%), pepper mild mottle virus (13.0 $\pm$ 4.4%), cucumber green mottle mosaic virus (11.5 $\pm$ 4.3%), tomato mosaic virus (5.0 $\pm$ 3.0%), tobacco mild green mosaic virus (2.6 $\pm$ 1.6%), tropical soda apple mosaic virus (1.8 $\pm$ 1.4%), tomato mottle mosaic virus (1.5 $\pm$ 1.4%), SARS-CoV-2 virus (0.9 $\pm$ 2.3%), crAssphage (0.4 $\pm$ 0.7%), and opuntia virus 2 (0.3 $\pm$ 0.6%) (Fig. 2). We were able to detect many more reads of SARS-CoV-2 with respiratory virus enrichment, as there were only 337 SARS-CoV-2 reads (an average proportional abundance of 0.0004%) in unenriched samples, while across enriched samples, we detected 124,135 SARS-CoV-2 reads. We note that IRV enrichment did not have a large impact on our ability to detect the most abundant viruses, rather it allowed us to detect the less abundant respiratory viruses relevant to public health.

The Illumina respiratory virus oligonucleotide panel enriches for 40 viruses (including SARS-CoV-2), so we were able to compare viral detection in 86 enriched and unenriched samples, along with two nonrespiratory viruses often detected in wastewater, norovirus and pepper mild mottle virus (PMMoV). In enriched libraries, we detected the presence of SARS-CoV-2 in 68 samples, human coronavirus (HCoV)-OC43 in 22 samples, HCoV-229E in one sample, influenza A in 10 samples, human adenoviruses in 15 samples, human bocaviruses in 13 samples, and noroviruses in 52 samples. In unenriched samples, we detected SARS-CoV-2 in 24 samples, HCoV-OC43 in two samples, and noroviruses in 59 samples. Likewise, we detected PMMoV in all samples regardless of enrichment (Fig. 3), indicating that sequencing was successful.

Wastewater viral ecology. We analyzed the Shannon indexes for unenriched samples and found that overall alpha diversity was significantly different between wastewater sampling sites ($F_{[6,87]} = 7.5$, $P < 0.001$) and then used Tukey's honestly significant difference (HSD) *post hoc* pairwise comparison testing to show that only NC-HTP, PL-HTP, PL-JWPCP, PL-OC, and PL-SJ alpha diversities were different from each other (adjusted P value [P_{adj}] < 0.05) (Fig. 4). We also compared the Bray-Curtis dissimilarities of the samples with Adonis and found that overall treatment plants' beta diversity values were significantly different ($R^2 = 0.41$, $P < 0.001$) (Fig. 4). Additionally, we used analysis of compositions of microbiomes (ANCOM) to test for differential abundance of viruses at greater than 0.0001 average relative abundance between treatment plants. This resulted in 16 viruses being differentially abundant between treatment plants ($W > 46$, $P_{adj} < 0.05$ each; (Fig. 4). In enriched samples, we compared the proportional
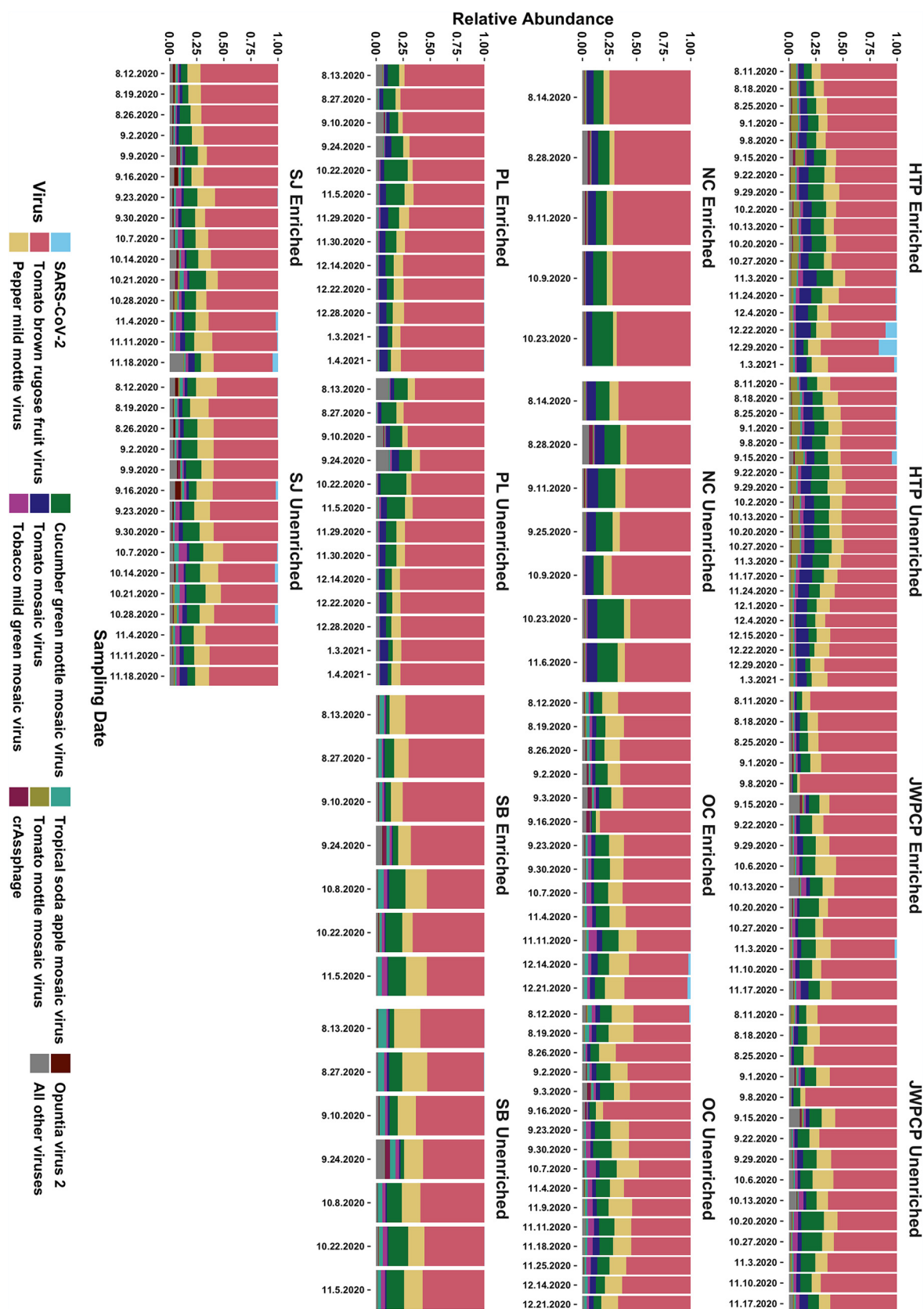
FIG 2 Stacked bar plot indicating the 10 most proportionally abundant viruses across all samples, with the relative abundance of SARS-CoV-2 being highest in HTP samples from 29 December 2020. Abbreviations denote WTP, "enriched" are samples enriched for respiratory viruses, and dates across the "*x*" axis signify sampling date.
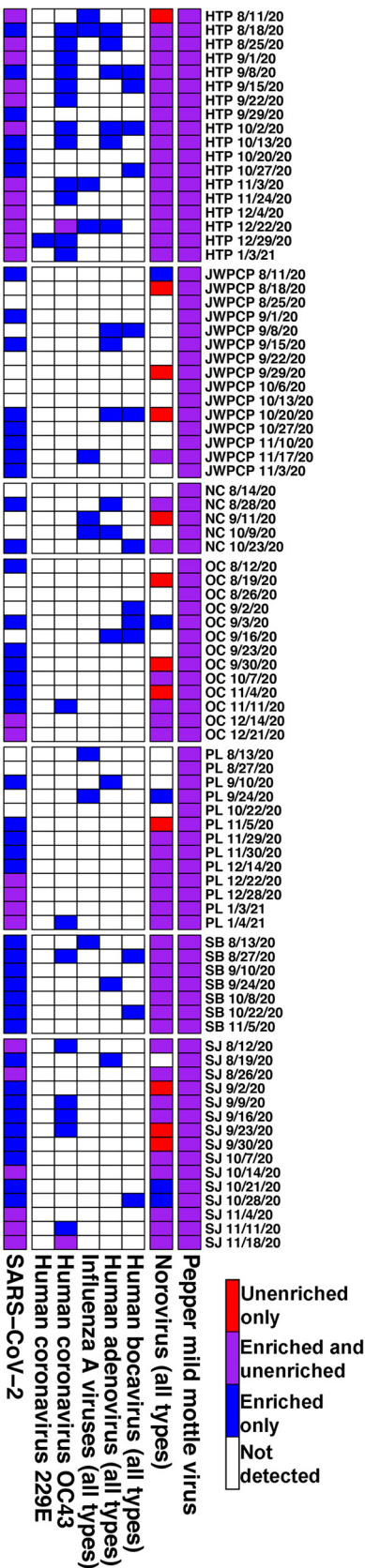
FIG 3 Heat map indicating the presence or absence of reads mapping to each listed virus. Heat color denotes if a virus was detected in unenriched samples, respiratory virus-enriched samples, both, or neither type of sample, and abbreviations signify WTP.
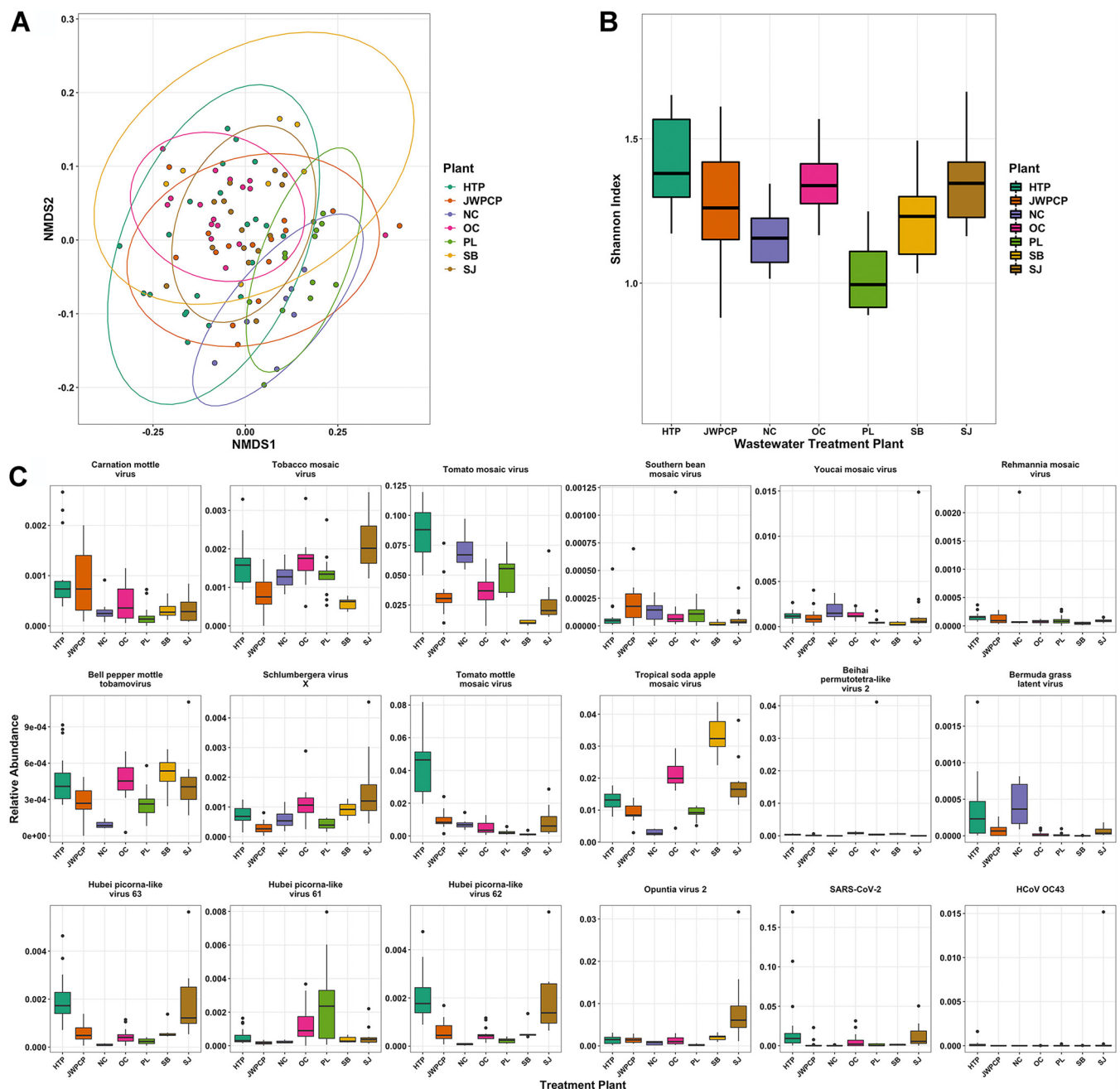
FIG 4 (A) Nonmetric multidimensional scaling (NMDS) ordination of the Bray-Curtis dissimilarities of all unenriched samples colored by WTP. Treatment plants were significantly different from each other ($R^2 = 0.41$, $P < 0.001$). (B) Box plots of the Shannon diversity indexes colored by WTP. Overall alpha diversity was significantly different ($F_{[6,87]} = 7.5$, $P < 0.001$). (C) Box plots of virus relative abundances that significantly differed between treatment plants as tested with ANCOM. The graphs on the bottom right (SARS-CoV-2 and HCoV OC43) represent enriched samples only.

abundances of human respiratory viruses at greater than 0.0001 average relative abundance and showed that SARS-CoV-2 and HCoV-OC43 were significantly different across treatment plants ($W > 46$, $P_{adj} < 0.05$ each) (Fig. 4).

As we sampled the treatment plants multiple times over our study (up to 145 days from first sample), we were able to study how the viromes changed longitudinally. We compared diversity measures over time with linear mixed effects (LMEs) using treatment plant as a random effect and show that both alpha and beta diversity remained stable across the sampling periods in unenriched samples (Shannon: $t = 0.21$, $P = 0.84$; Bray-Curtis: $t = -0.31$, $P = 0.76$). We also used LME on viruses present at greater than 0.0001 average relative abundance with treatment plant as a random effect and

showed that 19 viruses' relative abundances changed over time ($P_{adj} \leq 0.05$) (Fig. S1). We specifically were interested in SARS-CoV-2 in enriched samples and used LMEs with treatment plant as a random effect and found that the relative abundance of this virus increased between August 2020 and January 2021 ($t = 4.0$, $P < 0.001$) (Fig. S1).

Sequencing SARS-CoV-2 single-nucleotide variants in wastewater samples. We were interested in reads mapping specifically to the SARS-CoV-2 genome as a way to detect single-nucleotide variants (SNVs) in wastewater. Across the 68 enriched samples that had detectable SARS-CoV-2 reads, we obtained an average breadth of genomic coverage of 24.0% (range: 0.2% to 99.8%) per sample at an average sequencing depth of 8.1 reads per base (range: 0.002 to 177.6) (Fig. 5; Table S2). After masking the likely problematic nucleotide sites (as suggested by https://virological.org/t/masking-strategies-for-sars-cov-2-alignments/480/14, March 2021 update), we obtained 2,558 SNVs (2,002 unique) across all samples; however, due to the low breadth of coverage in many samples, many of these sites may be spurious or unresolved. After applying a more stringent cutoff of 50% breadth of coverage across the SARS-CoV-2 genome, we obtained 2,060 SNVs (1,656 unique) across 14 samples (Fig. 5; Table S3).

As we took samples at multiple time points per treatment plant, we were able to track the proportion of SNVs per nucleotide position over time, most notably through samples taken from the Hyperion (HTP) facility, likely due to higher viral load leading to higher sequencing depth. Within HTP, we plotted the proportion of SNVs (compared to the reference strain) over time at nucleotide positions obtained from at least three samples with greater than 50% breadth of genomic coverage. Three of the detected SNVs are apparently fixed in the viral population (sites 241, 14408, and 23403 are all 100% SNV), while the other 17 SNVs appear to vary widely over the sampling dates (overall average percent coefficient of variation [%CV] = 103%; range: 0% to 161%) with no apparent directionality (Fig. 6).

DISCUSSION

Our composite wastewater samples contained a diversity of RNA viruses (mainly plant-infecting viruses along with lower relative abundances of animal-infecting viruses) that differed between wastewater treatment plants (WTPs), supporting previous studies indicating that location affects the presence of viruses (13). Furthermore, several individual viruses varied over time while overall diversity remained unchanged, likely due to localized infections within WTP catchment areas (3, 9, 47). We detected several human-pathogenic viruses across all WTPs, supporting the hypothesis that wastewater-based epidemiology (WBE) has the potential to inform public health and researchers about viral presence and distribution without relying on standard health care practices (17). Through respiratory-virus-enriched library preparation and sequencing, we detected the presence of SARS-CoV-2 at every WTP, along with several SNVs across the SARS-CoV-2 genome. This suggests that WBE can reveal the pool of potential viral variants across large geographic areas, again without adding stress to health care systems, with the benefit that composite sampling can collect wastewater 24 h per day (45, 46, 48). Lastly, we show that SARS-CoV-2 viral load at WTPs generally correlates with county-level COVID-19 case counts, indicating that WBE can be useful in monitoring the severity and dynamics of disease spread (18, 19, 23–25, 49).

The vast majority of viruses present in our samples were plant viruses, mainly those infecting tomatoes, peppers, and cucumbers in the genus *Tobamovirus*. This result is consistent with other studies, which suggests that these viruses are diverse and widespread in wastewater, likely originating from agricultural runoff or human feces (3, 7, 50). Even though these viruses were ubiquitous throughout our samples, WTPs had significantly different overall viromes, indicating that there may be signatures of location and wastewater catchment throughout Southern California, which has been suggested from other sampling locations (13). Aside from overall diversity, several plant- or arthropod-infecting viruses were differentially abundant between WTPs, possibly due to differences in peoples' diet (and thus viral excretion), infected plant growth,
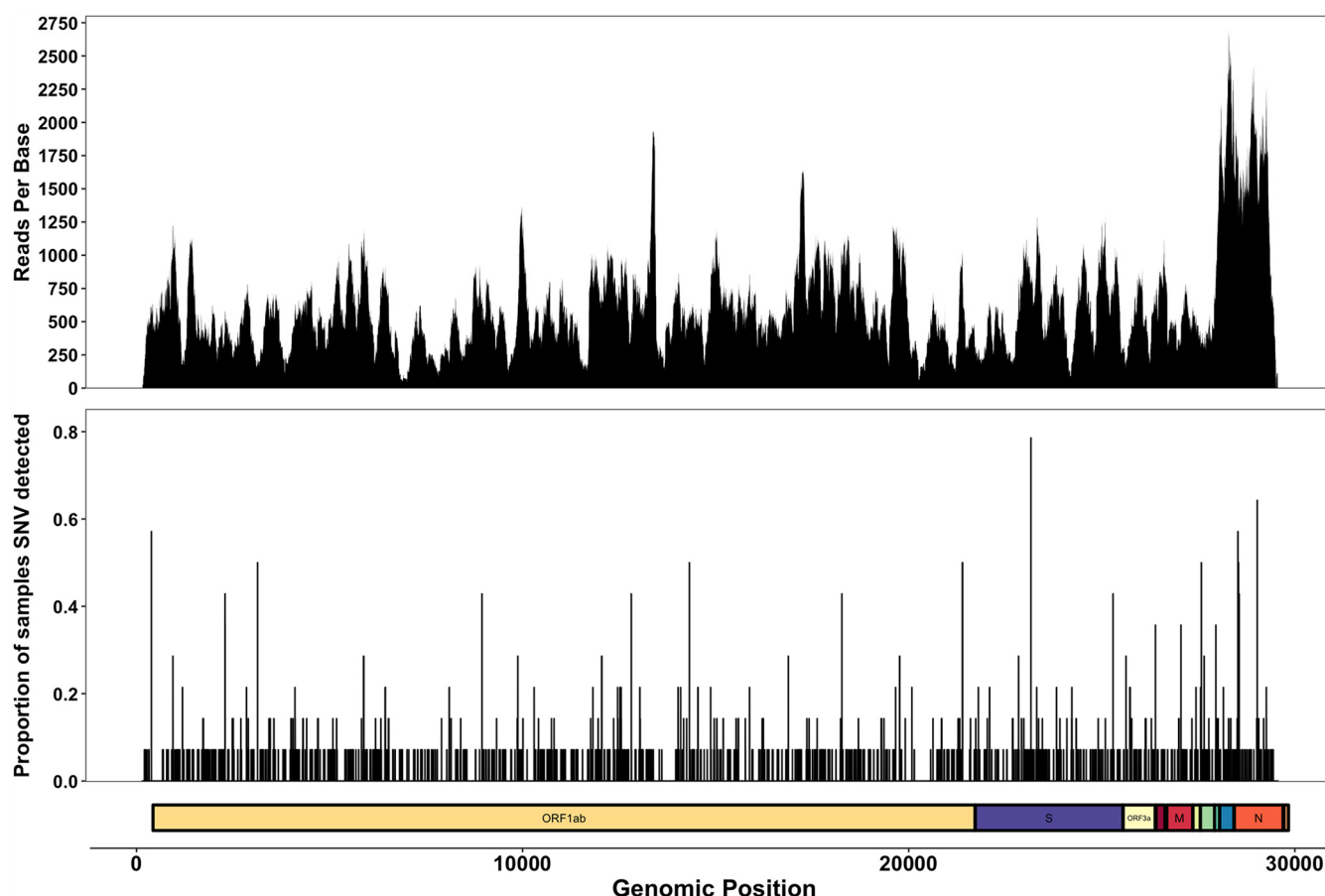
FIG 5 Multipanel figure showing sequence reads per base across all enriched samples, the position and proportion of SNVs across samples with >50% genomic breadth only, and an open reading frame (ORF) map of the SARS-CoV-2 genome.

and localized infections of arthropods (i.e., Hubei picorna-like viruses and Beihai per-mutotetra-like viruses) (3, 9, 47). Likewise, we found that several viruses' relative abun-dances varied over time, regardless of treatment plant, suggesting that there are infec-tion/subsidence dynamics or seasonal trends as previously suggested (7, 10, 49, 51). As the day-to-day relative abundance of viruses varied, we suggest that future wastewater surveys incorporate longitudinal and composite sampling to accurately capture viral diversity.

We detected several viruses pathogenic to humans across WTPs, including norovi-ruses, adenoviruses, bocaviruses, coronaviruses, and influenza A, which agrees with other studies, indicating that WBE is robust and applicable to multiple viruses (18–23). By applying metagenomic or metatranscriptomic sequencing to wastewater, we can simultaneously detect a diversity of viruses, potentially alerting public health to unknown or underreported infections or new viral strains (27, 28, 46, 52). As we pre-pared sequencing libraries using two methods (unenriched and Illumina respiratory vi-rus oligonucleotide panel-enriched shotgun metatranscriptomics), we could compare the effects of enrichment on virus detection and report that viral enrichment greatly improved our ability to detect influenza A and coronaviruses (especially SARS-CoV-2). We suggest that if researchers and the public health community are specifically inter-ested in respiratory viruses, enrichment and the proper wastewater concentration/extraction methods should be used for appropriate sensitivity and specificity to viruses of interest (23, 25, 34, 46, 53). Alongside sequencing, we also used ddPCR to quantify SARS-CoV-2 in wastewater and show that our results generally correlated well with county-reported 7-day rolling average COVID-19 cases, which agrees with previous studies (23, 24, 54). We note that SARS-CoV-2 viral load and case counts were not
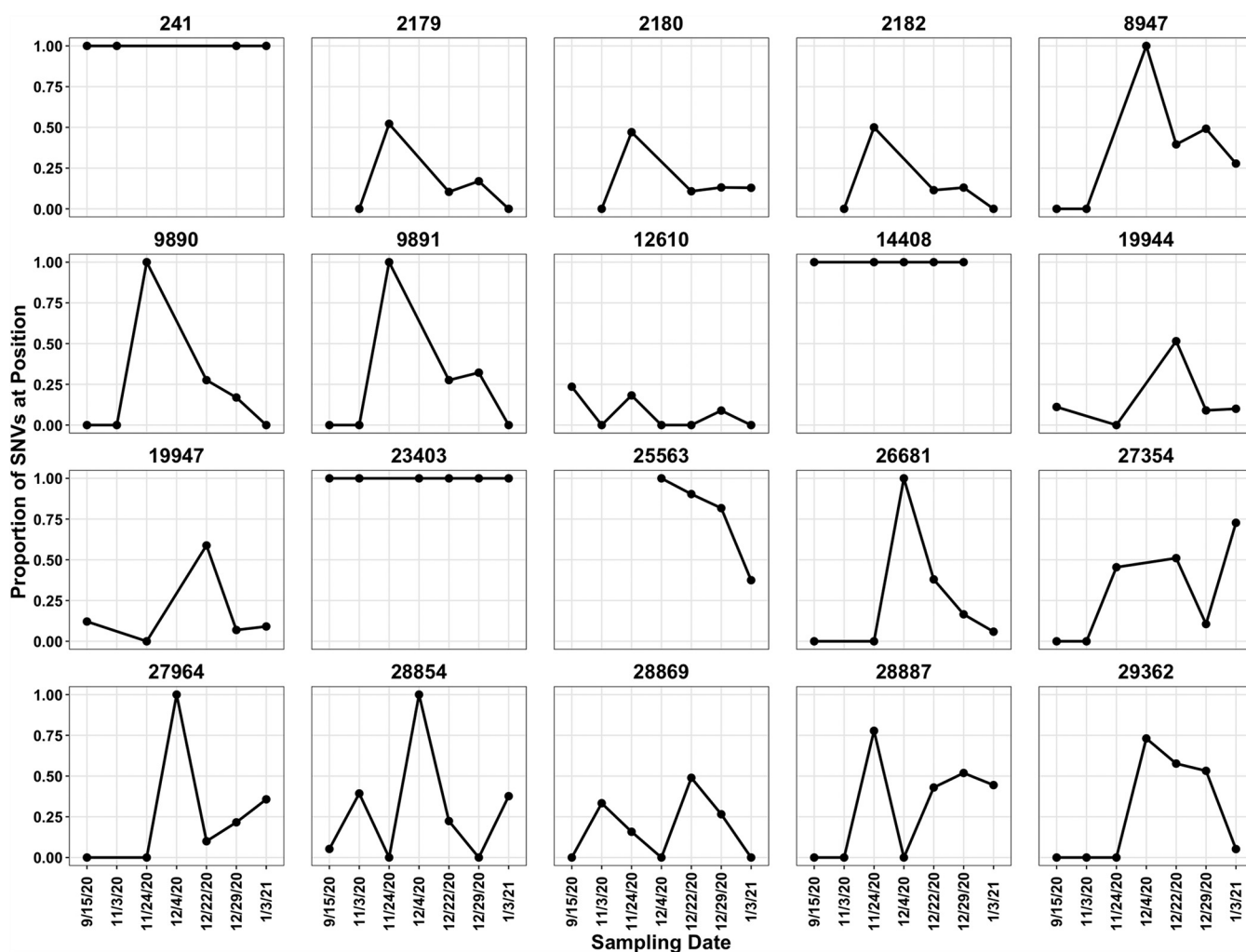
FIG 6 Line and dot plot of SNVs sequenced at least three times at the Hyperion WTP faceted by SARS-CoV-2 genomic position over time in samples with >50% sequencing breadth. "Proportion of SNVs" refers to the proportion of SNVs versus the SARS-CoV-2 reference strain (i.e., 1 SNV and 3 reference reads = 0.25 proportion).

significantly correlated at every WTP, indicating that there is likely unknown variability in the survival of RNA within wastewater, or that there may be variable influent flow or water quality affecting the viral load at specific WTPs (55, 56). We also note that our comparison is limited to county-level data rather than being a comparison of the areas contributing to the influent stream, and this could be obscuring the relationship-to-case data at some treatment plants.

Our respiratory virus-enriched metatranscriptomic sequencing detected the presence of SARS-CoV-2 at every WTP and in most of our samples, although we note that we could not accurately quantify viruses through sequencing. Instead, the power of metatranscriptomic sequencing lies in our ability to detect SNVs across the SARS-CoV-2 genome (46), and, depending on the sample, we often sequenced greater than 50% of the genome at appreciable read depth. For example, we detected the GSAID clade GH (lineage B.1*) markers 241C>T, 3037C>T, 14408C>T, 23403A>G, and 25563G>T at 100% prevalence where those regions of the genome were sequenced (39, 57, 58). Likewise, as we detected over 1,000 SARS-CoV-2 SNVs across our samples, we found many SNVs of putatively unknown function that have been detected in patient samples, such as 6285C>T and 9891C>T (found in, but does not solely define, variants B.1.525 and B.1.1.318, respectively) and 28854C>T and 28887C>T (40, 59, 60). We also detected many SNVs also found in wastewater samples from Northern California (46) and SNVs that, to the best of our knowledge, have yet to be sequenced (41, 46, 48). As

our sequencing data are not quantitative, our study suggests that sequencing waste-water is useful for SNV detection across wide catchment areas but is not useful for the true prevalence of SNVs (46). Furthermore, we recognize that wastewater likely represents a collection of different viruses' RNA rather than one intact virion, as SARS-CoV-2 may be relatively fragile in wastewater (55, 61). Lastly, we note that most SNVs came from samples with the highest number of SARS-CoV-2 reads and suggest sequencing samples deeply or using tiled amplicon-based approaches to obtain a useful breadth and depth of viral genomes for variant detection (46, 62).

Conclusion. Wastewater-based epidemiology has the potential to aid public health in monitoring the spread and severity of disease outbreaks. Our research contributes to the growing field of WBE by showing that Southern California wastewater harbors a diversity of RNA viruses and that these viral populations vary over time and between WTPs. Likewise, we are able to detect human viral pathogens without increasing the burden on local health care systems, further supporting the benefit of WBE. Through ddPCR and metatranscriptomic sequencing, we were able to measure the viral load of SARS-CoV-2 in wastewater and identify potentially novel SNVs, which may assist in monitoring viral evolution and the emergence of new variants. We suggest that future researchers use longitudinal metatranscriptomic sequencing on wastewater samples to further understand the spread of RNA viruses and how these viruses change over time.

MATERIALS AND METHODS

Sample collection. We collected 94 1-liter 24-h composite influent wastewater samples at seven WTPs across Southern California between August 2020 and January 2021 (Table S4 in the supplemental material). The samples were aliquoted into 50-ml tubes and stored at 4°C until sample processing. Note that extractions were performed independently for ddPCR quantification and viromic sequencing.

Wastewater sample processing for SARS-CoV-2 ddPCR quantification. We prepared influent samples for ddPCR following the method described in Steele et al. (63). Briefly, we first added bovine corona-virus (BoCoV) vaccine (Bovilis; Merck & Co, Kenilworth, NJ) to 20 ml of wastewater as a sample processing control to assess viral RNA recovery. We then added $MgCl_2$ to a final concentration of 25 mM and adjusted the pH to <3.5 with 20% HCl on a mixed cellulose ester membrane (type HA; Millipore, Bedford, MA) in replicates of six. We then transferred the HA filters to preloaded 2-ml ZR BashingBead lysis tubes (Zymo, Irvine, CA) and bead beat the samples with a BioSpec beadbeater (BioSpec Products, Bartlesville, OK) for 1 min. We then extracted total nucleic acids with a bioMérieux NucliSENS extraction kit with magnetic bead capture (bioMérieux, Durham, NC) following the manufacturer's protocol.

SARS-CoV-2 reverse transcription-droplet digital PCR quantification and correlation with countywide COVID-19 cases. We used one-step reverse transcription-droplet digital PCR (RT-ddPCR) to quantify the N1 region of the SARS-CoV-2 N gene with primer and probe sequences designed by the CDC (63, 64), and we quantified the bovine coronavirus using previously designed primers (63, 65). We set up RT-ddPCR reactions following the manufacturer's instructions on a Bio-Rad Qx200 (Bio-Rad, Hercules, CA). For all assays, a minimum of two reactions and a total of ≥20,000 droplets were generated per sample, and at least five no-template control (NTC) reactions and two positive-control reactions were run per 96-well plate as well as extraction-specific NTCs. Each sample was required to have a minimum of three positive droplets (66) to be included in further analyses. We also assessed RNA recovery using the BoCoV exogenous control, and samples with <3% recovery were excluded from further analyses.

We used the SARS-CoV-2 ddPCR quantification data and compared viral loads with county-level COVID-19 reported case data from the State of California Health and Human Services Agency (https://data.chhs.ca.gov/dataset/covid-19-time-series-metrics-by-county-and-state). We calculated rolling 7-day averages for COVID-19 cases in Los Angeles, Orange, and San Diego counties with the R package "zoo" v1.8-9 (67) and ran Pearson correlations between viral load and reported COVID-19 cases with the R package "Hmisc" v4.5 (68) at each time point where we had both data points.

Wastewater sample processing for metatranscriptomic sequencing. We followed a similar protocol as Crits-Christoph et al. (46) and Wu et al. (23) to concentrate viruses and extract RNA. We pasteurized 50 ml of wastewater in a 65°C water bath for 90 min and filtered the samples through a sterile 0.22-μm vacuum filter (VWR, Radnor, PA) to remove solids. We then concentrated the filtrate through ultracentrifugation at 3,000 × g with 10-kDa Amicon filters (MilliporeSigma, Burlington, MA) by successively centrifuging then discarding the flowthrough until the entire 50-ml sample was processed. This resulted in final volumes of less than 500 μl for each sample, which we then stored at −80°C until RNA extraction. We thawed the wastewater concentrate on ice, then used an Invitrogen PureLink RNA minikit with DNase (Invitrogen, Waltham, MA) to extract RNA by following the manufacturer's protocol, quantified the resulting RNA with an AccuBlue broad range RNA quantification kit (Biotium, Fremont, CA) on a DeNovix QFX fluorometer (DeNovix, Wilmington, DE) spectrophotometer, and stored the RNA at −80°C.

Next-generation sequencing library preparation. Sample library preparation and next-generation sequencing was performed by the University of California Irvine Genomics High Throughput Facility (GHTF). The GHTF used two separate library preparation strategies per sample: one library was prepared with the Illumina RNA prep with enrichment kit (Illumina, San Diego, CA), and the other library was

prepared using the same preparation kit with the addition of the Illumina respiratory virus oligonucleo-tide panel to enrich for human respiratory viruses. The GHTF then sequenced the paired-end libraries ei-ther 2 × 100 bp or 2 × 150 bp (Table S4) on an Illumina NovaSeq 6000 with an S4 300 cycle kit and sent the data as demultiplexed FASTQ files.

Bioinformatics and sequence data processing. We used the University of California Irvine High Performance Community Computing Cluster (HPC3) for all data processing on the provided FASTQ files. We removed primers, adapter sequences, and low-quality bases with "bbduk" in the BBTools software package v38.87 (69) and removed PCR duplicates with BBTools "dedupe." We downloaded the NCBI Virus RefSeq Genome database (January 2021), used Bowtie2 v2.4.1 (70) to build an index of the viral genomes, and mapped the cleaned FASTQ reads against this index. We used "Samtools" v1.10 (71) to convert the resulting SAM files to sorted BAM files and finally used "inStrain" v1.3.1 (72) to calculate all viral abundances and profile viral variants on read pairs with >90% average nucleotide identity to their respective reference genomes. Lastly, we used "mosdepth" v0.3.1 (73) and "bedtools" v2.30.0 (74) to cal-culate genome coverage and breadth across all SARS-CoV-2 sequences.

For community diversity analyses, we tabulated viral abundances and normalized the reads into within-sample relative abundances in R v4.0.4 (75). We used this table to generate Shannon diversity indices and Bray-Curtis dissimilarity matrices with the R package "vegan" v2.5-7 (76). We also ran Adonis permutational multivariate analysis of variance (PERMANOVA) tests on the distance matrices and per-formed nonmetric multidimensional scaling on the data, compared the proportional abundances of viruses with ANCOM v2.1 (77), and compared the relative abundances of viruses over time with "lmerTest" v3.1-3 using WTP as a random effect (78). Lastly, we plotted all figures with "ggplot2" v3.3.3 (79), "gggenes" v0.4.1 (80), or "pheatmap" v1.0.12 (81).

Data availability. Raw sequencing data have been deposited on the NCBI Sequence Read Archive under accession number PRJNA729801, and representative code can be found at https://github.com/jasonarothman/wastewater_viromics_sarscov2.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, XLSX file, 0.01 MB.
SUPPLEMENTAL FILE 2, XLSX file, 0.02 MB.
SUPPLEMENTAL FILE 3, XLSX file, 0.2 MB.
SUPPLEMENTAL FILE 4, XLSX file, 0.02 MB.
SUPPLEMENTAL FILE 5, PDF file, 0.6 MB.

REFERENCES

1. Newton RJ, McClary JS. 2019. The flux and impact of wastewater infra-structure microorganisms on human and ecosystem health. Curr Opin Biotechnol 57:145–150. https://doi.org/10.1016/j.copbio.2019.03.015.

2. Wu L, Ning D, Zhang B, Li Y, Zhang P, Shan X, Zhang Q, Brown MR, Li Z, Van Nostrand JD, Ling F, Xiao N, Zhang Y, Vierheilig J, Wells GF, Yang Y, Deng Y, Tu Q, Wang A, Global Water Microbiome Consortium, Zhang T, He Z, Keller J, Nielsen PH, Alvarez PJJ, Criddle CS, Wagner M, Tiedje JM, He Q, Curtis TP, Stahl DA, Alvarez-Cohen L, Rittmann BE, Wen X, Zhou J. 2019. Global diversity and biogeography of bacterial communities in waste-water treatment plants. Nat Microbiol 4:1183–1195. https://doi.org/10.1038/s41564-019-0426-5.

3. Symonds EM, Griffin DW, Breitbart M. 2009. Eukaryotic viruses in wastewater samples from the United States. Appl Environ Microbiol 75:1402–1409. https://doi.org/10.1128/AEM.01899-08.

4. McLellan SL, Huse SM, Mueller-Spitz SR, Andreishcheva EN, Sogin ML. 2010. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. Environ Microbiol 12:378–392. https://doi.org/10.1111/j.1462-2920.2009.02075.x.

5. Karkman A, Do TT, Walsh F, Virta MPJ. 2018. Antibiotic-resistance genes in waste water. Trends Microbiol 26:220–228. https://doi.org/10.1016/j.tim.2017.09.005.

6. Edwards RA, Vega AA, Norman HM, Ohaeri M, Levi K, Dinsdale EA, Cinek O, Aziz RK, McNair K, Barr JJ, Bibby K, Brouns SJJ, Cazares A, de Jonge PA, Desnues C, Díaz Muñoz SL, Fineran PC, Kurilshikov A, Lavigne R, Mazankova K, McCarthy DT, Nobrega FL, Reyes Muñoz A, Tapia G, Trefault N, Tyakht AV, Vinuesa P, Wagemans J, Zhernakova A, Aarestrup FM, Ahmadov G, Alassaf A, Anton J, Asangba A, Billings EK, Cantu VA, Carlton JM, Cazares D, Cho G-S, Condeff T, Cortés P, Cranfield M, Cuevas DA, De la

Iglesia R, Decewicz P, Doane MP, Dominy NJ, Dziewit L, Elwasila BM, Eren AM, Franz C, Fu J, Garcia-Aljaro C, et al. 2019. Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. Nat Microbiol 4:1727–1736. https://doi.org/10.1038/s41564-019-0494-6.

7. Cantalupo PG, Calgua B, Zhao G, Hundesa A, Wier AD, Katz JP, Grabe M, Hendrix RW, Girones R, Wang D, Pipas JM. 2011. Raw sewage harbors diverse viral populations. mBio 2:e00180-11. https://doi.org/10.1128/mBio.00180-11.

8. Rosario K, Nilsson C, Lim YW, Ruan Y, Breitbart M. 2009. Metagenomic analysis of viruses in reclaimed water. Environ Microbiol 11:2806–2820. https://doi.org/10.1111/j.1462-2920.2009.01964.x.

9. Kitajima M, Sassi HP, Torrey JR. 2018. Pepper mild mottle virus as a water quality indicator. NPJ Clean Water 1:19. https://doi.org/10.1038/s41545-018-0019-5.

10. Eftim SE, Hong T, Soller J, Boehm A, Warren I, Ichida A, Nappier SP. 2017. Occurrence of norovirus in raw sewage—a systematic literature review and meta-analysis. Water Res 111:366–374. https://doi.org/10.1016/j.watres.2017.01.017.

11. Martínez-Puchol S, Rusiñol M, Fernández-Cassi X, Timoneda N, Itarte M, Andrés C, Antón A, Abril JF, Girones R, Bofill-Mas S. 2020. Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens. Sci Total Environ 713:136604. https://doi.org/10.1016/j.scitotenv.2020.136604.

12. Ng TFF, Marine R, Wang C, Simmonds P, Kapusinszky B, Bodhidatta L, Oderinde BS, Wommack KE, Delwart E. 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J Virol 86:12161–12175. https://doi.org/10.1128/JVI.00869-12.

13. Bibby K, Peccia J. 2013. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. Environ Sci Technol 47:1945–1951. https://doi.org/10.1021/es305181x.

14. Ferrera I, Sánchez O. 2016. Insights into microbial diversity in wastewater treatment systems: how far have we come? Biotechnol Adv 34:790–802. https://doi.org/10.1016/j.biotechadv.2016.04.003.

15. Bivins A, North D, Ahmad A, Ahmed W, Alm E, Been F, Bhattacharya P, Bijlsma L, Boehm AB, Brown J, Buttiglieri G, Calabro V, Carducci A, Castiglioni S, Cetecioglu Gurol Z, Chakraborty S, Costa F, Curcio S, De Los Reyes FL, Delgado Vela J, Farkas K, Fernandez-Casi X, Gerba C, Gerrity D, Girones R, Gonzalez R, Haramoto E, Harris A, Holden PA, Islam MT, Jones DL, Kasprzyk-Hordern B, Kitajima M, Kotlarz N, Kumar M, Kuroda K, La Rosa G, Malpei F, Mautus M, McLellan SL, Medema G, Meschke JS, Mueller J, Newton RJ, Nilsson D, Noble RT, Van Nuijs A, Peccia J, Perkins TA, Pickering AJ, et al. 2020. Wastewater-based epidemiology: global collaborative to maximize contributions in the fight against COVID-19. Environ Sci Technol 54:7754–7757. https://doi.org/10.1021/acs.est.0c02388.

16. Carrasco-Hernandez R, Jácome R, López Vidal Y, Ponce de León S. 2017. Are RNA viruses candidate agents for the next global pandemic? A review. ILAR J 58:343–358. https://doi.org/10.1093/ilar/ilx026.

17. Sims N, Kasprzyk-Hordern B. 2020. Future perspectives of wastewater-based epidemiology: monitoring infectious disease spread and resistance to the community level. Environ Int 139:105689. https://doi.org/10.1016/j.envint.2020.105689.

18. Hellmér M, Paxéus N, Magnius L, Enache L, Arnholm B, Johansson A, Bergström T, Norder H. 2014. Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. Appl Environ Microbiol 80:6771–6781. https://doi.org/10.1128/AEM.01981-14.

19. Manor Y, Handsher R, Halmut T, Neuman M, Bobrov A, Rudich H, Vonsover A, Shulman L, Kew O, Mendelson E. 1999. Detection of poliovirus circulation by environmental surveillance in the absence of clinical cases in Israel and the Palestinian authority. J Clin Microbiol 37:1670–1675. https://doi.org/10.1128/JCM.37.6.1670-1675.1999.

20. Blomqvist S, El Bassioni L, El Maamoon Nasr EM, Paananen A, Kaijalainen S, Asghar H, de Gourville E, Roivainen M. 2012. Detection of imported wild polioviruses and vaccine-derived polioviruses by environmental surveillance in Egypt. Appl Environ Microbiol 78:5406–5409. https://doi.org/10.1128/AEM.00491-12.

21. Heijnen L, Medema G. 2011. Surveillance of influenza A and the pandemic influenza A (H1N1) 2009 in sewage and surface water in the Netherlands. J Water Health 9:434–442. https://doi.org/10.2166/wh.2011.019.

22. Wang XW, Li J, Guo T, Zhen B, Kong Q, Yi B, Li Z, Song N, Jin M, Xiao W, Zhu X, Gu C, Yin J, Wei W, Yao W, Liu C, Li J, Ou G, Wang M, Fang T, Wang G, Qiu Y, Wu H, Chao F, Li J. 2005. Concentration and detection of SARS coronavirus in sewage from Xiao Tang Shan Hospital and the 309th Hospital of the Chinese People's Liberation Army. Water Sci Technol 52:213–221. https://doi.org/10.2166/wst.2005.0266.

23. Wu F, Zhang J, Xiao A, Gu X, Lee WL, Armas F, Kauffman K, Hanage W, Matus M, Ghaeli N, Endo N, Duvallet C, Poyet M, Moniz K, Washburne AD, Erickson TB, Chai PR, Thompson J, Alm EJ. 2020. SARS-CoV-2 titers in wastewater are higher than expected from clinically confirmed cases. mSystems 5:e00614-20. https://doi.org/10.1128/mSystems.00614-20.

24. Karthikeyan S, Ronquillo N, Belda-Ferre P, Alvarado D, Javidi T, Longhurst CA, Knight R. 2021. High-throughput wastewater SARS-CoV-2 detection enables forecasting of community infection dynamics in San Diego County. mSystems 6:e00045-21. https://doi.org/10.1128/mSystems.00045-21.

25. Peccia J, Zulli A, Brackney DE, Grubaugh ND, Kaplan EH, Casanovas-Massana A, Ko AI, Malik AA, Wang D, Wang M, Warren JL, Weinberger DM, Arnold W, Omer SB. 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. Nat Biotechnol 38:1164–1167. https://doi.org/10.1038/s41587-020-0684-z.

26. Kopel E, Kaliner E, Grotto I. 2014. Lessons from a public health emergency—importation of wild poliovirus to Israel. N Engl J Med 371:981–983. https://doi.org/10.1056/NEJMp1406250.

27. Gibbons CL, Mangen M-JJ, Plass D, Havelaar AH, Brooke RJ, Kramarz P, Peterson KL, Stuurman AL, Cassini A, Fèvre EM, Kretzschmar MEE, Burden of Communicable diseases in Europe (BCoDE) consortium. 2014. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. BMC Public Health 14:147. https://doi.org/10.1186/1471-2458-14-147.

28. Silverman JD, Hupert N, Washburne AD. 2020. Using influenza surveillance networks to estimate state-specific prevalence of SARS-CoV-2 in the United States. Sci Transl Med 12:eabc1126. https://doi.org/10.1126/scitranslmed.abc1126.

29. Medema G, Heijnen L, Elsinga G, Italiaander R, Brouwer A. 2020. 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 7:511–516. https://doi.org/10.1021/acs.estlett.0c00357.

30. Wigginton KR, Ye Y, Ellenberg RM. 2015. Emerging investigators series: the source and fate of pandemic viruses in the urban water cycle. Environ Sci Water Res Technol 1:735–746. https://doi.org/10.1039/C5EW00125K.

31. Dong E, Du H, Gardner L. 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 20:533–534. https://doi.org/10.1016/S1473-3099(20)30120-1.

32. Chan JF-W, Kok K-H, Zhu Z, Chu H, To KK-W, Yuan S, Yuen K-Y. 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect 9:221–236. https://doi.org/10.1080/22221751.2020.1719902.

33. World Health Organization. 9 January 2020. WHO Statement regarding cluster of pneumonia cases in Wuhan, China. https://www.who.int/china/news/detail/09-01-2020-who-statement-regarding-cluster-of-pneumonia-cases-in-wuhan-china. Accessed 15 March 2021.

34. Rothman JA, Loveless TB, Griffith ML, Steele JA, Griffith JF, Whiteson KL. 2020. Metagenomics of wastewater influent from Southern California wastewater treatment facilities in the era of COVID-19. Microbiol Resour Announc 9:e00907-20. https://doi.org/10.1128/MRA.00907-20.

35. Prado T, Fumian TM, Mannarino CF, Resende PC, Motta FC, Eppinghaus ALF, Chagas do Vale VH, Braz RMS, de Andrade J da SR, Maranhão AG, Miagostovich MP. 2021. Wastewater-based epidemiology as a useful tool to track SARS-CoV-2 and support public health policies at municipal level in Brazil. Water Res 191:116810. https://doi.org/10.1016/j.watres.2021.116810.

36. Wu Y, Guo C, Tang L, Hong Z, Zhou J, Dong X, Yin H, Xiao Q, Tang Y, Qu X, Kuang L, Fang X, Mishra N, Lu J, Shan H, Jiang G, Huang X. 2020. Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. Lancet Gastroenterol Hepatol 5:434–435. https://doi.org/10.1016/S2468-1253(20)30083-2.

37. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva TI, Sheffield COVID-19 Genomics Group, McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO, Montefiori DC. 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 182:812–827. https://doi.org/10.1016/j.cell.2020.06.043.

38. Darby AC, Hiscox JA. 2021. Covid-19: variants and vaccination. BMJ 372:n771. https://doi.org/10.1136/bmj.n771.

39. Elbe S, Buckland-Merrett G. 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall 1:33–46. https://doi.org/10.1002/gch2.1018.

40. Badua CLDC, Baldo KAT, Medina PMB. 2021. Genomic and proteomic mutation landscapes of SARS-CoV-2. J Med Virol 93:1702–1721. https://doi.org/10.1002/jmv.26548.

41. Fang S, Li K, Shen J, Liu S, Liu J, Yang L, Hu C-D, Wan J. 2021. GESS: a database of global evaluation of SARS-CoV-2/hCoV-19 sequences. Nucleic Acids Res 49:D706–D714. https://doi.org/10.1093/nar/gkaa808.

42. Rouchka EC, Chariker JH, Chung D. 2020. Variant analysis of 1,040 SARS-CoV-2 genomes. PLoS One 15:e0241535. https://doi.org/10.1371/journal.pone.0241535.

43. Leite H, Lindsay C, Kumar M. 2021. COVID-19 outbreak: implications on healthcare operations. TQM J 33:247–256. https://doi.org/10.1108/TQM-05-2020-0111.

44. IHME COVID-19 Forecasting Team, Reiner RC, Jr, Barber RM, Collins JK, Zheng P, Adolph C, Albright J, Antony CM, Aravkin AY, Bachmeier SD, Bang-Jensen B, Bannick MS, Bloom S, Carter A, Castro E, Causey K, Chakrabarti S, Charlson FJ, Cogen RM, Combs E, Dai X, Dangel WJ, Earl L, Ewald SB, Ezalarab M, Ferrari AJ, Flaxman A, Frostad JJ, Fullman N, Gakidou E, Gallagher J, Glenn SD, Goosmann EA, He J, Henry NJ, Hulland EN, Hurst B, Johanns C, Kendrick PJ, Khemani A, Larson SL, Lazzar-Atwood A, LeGrand KE, Lescinsky H, Lindstrom A, Linebarger E, Lozano R, Ma R, Månsson J, Magistro B, Herrera AMM, Marczak LB, Miller-Petrie MK, Mokdad AH, et al. 2021. Modeling COVID-19 scenarios for the United States. Nat Med 27:94–105. https://doi.org/10.1038/s41591-020-1132-9.

45. Nemudryi A, Nemudraia A, Wiegand T, Surya K, Buyukyoruk M, Cicha C, Vanderwood KK, Wilkinson R, Wiedenheft B. 2020. Temporal detection and phylogenetic assessment of SARS-CoV-2 in municipal wastewater. Cell Rep Med 1:100098. https://doi.org/10.1016/j.xcrm.2020.100098.

46. Crits-Christoph A, Kantor RS, Olm MR, Whitney ON, Al-Shayeb B, Lou YC, Flamholz A, Kennedy LC, Greenwald H, Hinkle A, Hetzel J, Spitzer S, Koble J, Tan A, Hyde F, Schroth G, Kuersten S, Banfield JF, Nelson KL. 2021. Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 variants. mBio 12:e02703-20. https://doi.org/10.1128/mBio.02703-20.

47. Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X, Qin X-C, Li J, Cao J-P, Eden J-S, Buchmann J, Wang W, Xu J, Holmes EC, Zhang Y-Z. 2016. Redefining the invertebrate RNA virosphere. Nature 540:539–543. https://doi.org/10.1038/nature20167.

48. Fontenele RS, Kraberger S, Hadfield J, Driver EM, Bowes D, Holland LA, Faleye TOC, Adhikari S, Kumar R, Inchausti R, Holmes WK, Deitrick S, Brown P, Duty D, Smith T, Bhatnagar A, Yeager RA, Holm RH, Hoogesteijn von Reitzenstein N, Wheeler E, Dixon K, Constantine T, Wilson MA, Lim ES, Jiang X, Halden RU, Scotch M, Varsani A. 2021. High-throughput sequencing of SARS-CoV-2 in wastewater provides insights into circulating variants. medRxiv https://doi.org/10.1101/2021.01.22.21250320.

49. Kazama S, Masago Y, Tohma K, Souma N, Imagawa T, Suzuki A, Liu X, Saito M, Oshitani H, Omura T. 2016. Temporal dynamics of norovirus determined through monitoring of municipal wastewater by pyrosequencing and virological surveillance of gastroenteritis cases. Water Res 92:244–253. https://doi.org/10.1016/j.watres.2015.10.024.

50. Bačnik K, Kutnjak D, Pecman A, Mehle N, Tušek Žnidarič M, Gutiérrez Aguirre I, Ravnikar M. 2020. Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment. Water Res 177:115628. https://doi.org/10.1016/j.watres.2020.115628.

51. Brinkman NE, Fout GS, Keely SP. 2017. Retrospective surveillance of wastewater to examine seasonal dynamics of enterovirus infections. mSphere 2:e00099-17. https://doi.org/10.1128/mSphere.00099-17.

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

53. Philo SE, Keim EK, Swanstrom R, Ong AQW, Burnor EA, Kossik AL, Harrison JC, Demeke BA, Zhou NA, Beck NK, Shirai JH, Meschke JS. 2021. A comparison of SARS-CoV-2 wastewater concentration methods for environmental surveillance. Sci Total Environ 760:144215. https://doi.org/10.1016/j.scitotenv.2020.144215.

54. Wu F, Xiao A, Zhang J, Moniz K, Endo N, Armas F, Bonneau R, Brown MA, Bushman M, Chai PR, Duvallet C, Erickson TB, Foppe K, Ghaeli N, Gu X, Hanage WP, Huang KH, Lee WL, Matus M, McElroy KA, Nagler J, Rhode SF, Santillana M, Tucker JA, Wuertz S, Zhao S, Thompson J, Alm EJ. 2020. SARS-CoV-2 titers in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. medRxiv https://doi.org/10.1101/2020.06.15.20117747.

55. Bivins A, Greaves J, Fischer R, Yinda KC, Ahmed W, Kitajima M, Munster VJ, Bibby K. 2020. Persistence of SARS-CoV-2 in water and wastewater. Environ Sci Technol Lett 7:937–942. https://doi.org/10.1021/acs.estlett.0c00730.

56. Bivins A, North D, Wu Z, Shaffer M, Ahmed W, Bibby K. 2021. Within-day variability of SARS-CoV-2 RNA in municipal wastewater influent during periods of varying COVID-19 prevalence and positivity. ACS ES T Water 9:2097–2108. https://doi.org/10.1021/acsestwater.1c00178.

57. Alm E, Broberg EK, Connor T, Hodcroft EB, Komissarov AB, Maurer-Stroh S, Melidou A, Neher RA, O'Toole Á, Pereyaslov D, WHO European Region sequencing laboratories and GISAID EpiCoV group, WHO European Region sequencing laboratories and GISAID EpiCoV group. 2020. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. Euro Surveill 25:2001410. https://doi.org/10.2807/1560-7917.ES.2020.25.32.2001410.

58. Wang R, Chen J, Gao K, Hozumi Y, Yin C, Wei G-W. 2021. Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants. Commun Biol 4:228. https://doi.org/10.1038/s42003-021-01754-6.

59. Gupta A, Sabarinathan R, Bala P, Donipadi V, Vashisht D, Katika MR, Kandakatla M, Mitra D, Dalal A, Bashyam MD. 2020. Mutational landscape and dominant lineages in the SARS-CoV-2 infections in the state of Telangana, India. medRxiv https://doi.org/10.1101/2020.08.24.20180810.

60. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, Du Plessis L, Pybus OG. 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 5:1403–1407. https://doi.org/10.1038/s41564-020-0770-5.

61. Amoah ID, Kumari S, Bux F. 2020. Coronaviruses in wastewater processes: source, fate and potential risks. Environ Int 143:105962. https://doi.org/10.1016/j.envint.2020.105962.

62. Liu T, Chen Z, Chen W, Chen X, Hosseini M, Yang Z, Li J, Ho D, Turay D, Gheorghe C, Jones W, Wang C. 2021. A benchmarking study of SARS-CoV-2 whole-genome sequencing protocols using COVID-19 patient samples. iScience 24:102892. https://doi.org/10.1016/j.isci.2021.102892.

63. Steele JA, Zimmer-Faust AG, Griffith JF, Weisberg SB. 2021. Sources of variability in methods for processing, storing, and concentrating SARS-CoV-2 in influent from urban wastewater treatment plants. medRxiv https://doi.org/10.1101/2021.06.16.21259063.

64. Lu X, Wang L, Sakthivel SK, Whitaker B, Murray J, Kamili S, Lynch B, Malapati L, Burke SA, Harcourt J, Tamin A, Thornburg NJ, Villanueva JM, Lindstrom S. 2020. US CDC real-time reverse transcription PCR panel for detection of severe acute respiratory syndrome coronavirus 2. Emerg Infect Dis 26:1654–1665. https://doi.org/10.3201/eid2608.201246.

65. Decaro N, Elia G, Campolo M, Desario C, Mari V, Radogna A, Colaianni ML, Cirone F, Tempesta M, Buonavoglia C. 2008. Detection of bovine coronavirus using a TaqMan-based real-time RT-PCR assay. J Virol Methods 151:167–171. https://doi.org/10.1016/j.jviromet.2008.05.016.

66. Cao Y, Raith MR, Griffith JF. 2015. Droplet digital PCR for simultaneous quantification of general and human-associated fecal indicators for water quality assessment. Water Res 70:337–349. https://doi.org/10.1016/j.watres.2014.12.008.

67. Zeileis A, Grothendieck G. 2005. zoo: S3 infrastructure for regular and irregular time series. J Stat Softw 14:1–27. https://doi.org/10.18637/jss.v014.i06.

68. Harrell FE. 2019. Hmisc: Harrell miscellaneous. https://hbiostat.org/R/Hmisc/. Accessed 1 April 2021.

69. Bushnell B. 2014. BBTools software package. https://sourceforge.net/projects/bbmap/. Accessed 1 April 2021.

70. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.

71. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

72. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF. 2021. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nat Biotechnol 39:727–736. https://doi.org/10.1038/s41587-020-00797-0.

73. Pedersen BS, Quinlan AR. 2018. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics 34:867–868. https://doi.org/10.1093/bioinformatics/btx699.

74. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. https://doi.org/10.1093/bioinformatics/btq033.

75. R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

76. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2017 vegan: community ecology package. https://github.com/vegandevs/vegan. Accessed on 1 April 2021.

77. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. 2015. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis 26:27663. https://doi.org/10.3402/mehd.v26.27663.

78. Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. lmerTest package: tests in linear mixed effects models. J Stat Softw 82:1–26. https://doi.org/10.18637/jss.v082.i13.

79. Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, NY.

80. Wilkins D. 2020. gggenes: draw gene arrow maps in "ggplot2." https://wilkox.org/gggenes/. Accessed on 1 April 2021.

81. Kolde R. 2019. pheatmap: pretty heatmaps. https://cran.r-project.org/web/packages/pheatmap/index.html. Accessed on 1 April 2021.