



Article

Longitudinal Sequencing and Variant Detection of SARS-CoV-2 across Southern California Wastewater

Jason A. Rothman ^{1,*}, Andrew Saghir ¹, Amity G. Zimmer-Faust ², Kylie Langlois ², Kayla Raygoza ^{1,2}, Joshua A. Steele ², John F. Griffith ² and Katrine L. Whiteson ^{1,*}

¹ Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92697, USA

² Southern California Coastal Water Research Project, University of California, Costa Mesa, CA 92697, USA; joshuas@sccwrp.org (J.A.S.)

* Correspondence: rothmanj@uci.edu (J.A.R.); katrine@uci.edu (K.L.W.); Tel.: +1-(949)-824-3509 (J.A.R.); +1-(949)-824-9032 (K.L.W.)

Abstract: Wastewater-based epidemiology (WBE) is useful for detecting pathogen prevalence and may serve to effectively monitor diseases across broad scales. WBE has been used throughout the COVID-19 pandemic to track disease burden through quantifying SARS-CoV-2 RNA present in wastewater. Aside from case load estimation, WBE is being used to assay viral genomic diversity and emerging potential SARS-CoV-2 variants. Here, we present a study in which we sequenced RNA extracted from sewage influent obtained from eight wastewater treatment plants representing 16 million people in Southern California from April 2020 to August 2021. We sequenced SARS-CoV-2 with two methods: Illumina Respiratory Virus-Enriched metatranscriptomic sequencing (N = 269), and QIAseq SARS-CoV-2-tiled amplicon sequencing (N = 95). We classified SARS-CoV-2 reads into lineages and sublineages that approximated named variants and identified single nucleotide variants (SNVs), of which many are putatively novel SNVs and SNVs of unknown potential function and prevalence. Through our retrospective study, we also show that several SARS-CoV-2 sublineages were detected in wastewater before clinical detection, which may assist in the prediction of future variants of concern. Lastly, we show that sublineage diversity was similar across Southern California and that diversity changed over time, indicating that WBE is effective across megaregions. As the COVID-19 pandemic moves to new phases, and SARS-CoV-2 variants emerge, monitoring wastewater is important to understand local- and population-level dynamics of the virus. These results will aid in our ability to monitor the evolutionary potential of SARS-CoV-2 and help understand circulating SNVs to further combat COVID-19.

Keywords: COVID-19; SARS-CoV-2; wastewater-based epidemiology; metatranscriptomics



Citation: Rothman, J.A.; Saghir, A.; Zimmer-Faust, A.G.; Langlois, K.; Raygoza, K.; Steele, J.A.; Griffith, J.F.; Whiteson, K.L. Longitudinal Sequencing and Variant Detection of SARS-CoV-2 across Southern California Wastewater. *Appl. Microbiol.* **2024**, *4*, 635–649. <https://doi.org/10.3390/applmicrobiol4020044>

Academic Editor: Ian Connerton

Received: 29 February 2024

Revised: 27 March 2024

Accepted: 28 March 2024

Published: 29 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The COVID-19 pandemic has had a profound impact on the human population, causing over 700 million cases of disease and more than 7 million human deaths worldwide [1]. Caused by the emergence of the +ssRNA “severe acute respiratory syndrome coronavirus 2” [2], COVID-19 has caused public health to react and respond in novel ways to track the spread of the disease [3–5]. One of these unexpected responses has been through the use of wastewater-based epidemiology (WBE) to monitor SARS-CoV-2 viral loads in wastewater throughout the COVID-19 pandemic [4,6,7]. Having been used for approximately 80 years (although only formally adopted by the WHO in 2003) to monitor polio, WBE is known to be an effective surveillance method for a variety of diseases [8]. As part of a worldwide effort to combat COVID-19, a massive assemblage of epidemiologists has developed methods and analyses for the examination of SARS-CoV-2 viral material in wastewater to track the spread and approximate cases of COVID-19 [4,9–11]. While direct sampling from patients is the most definitive method of COVID-19 diagnosis [12,13], it has been shown that clinical testing has probably undercounted the true number of cases [3,14–16]. This inaccuracy in

case counts is likely due to a combination of supply issues, inability or reluctance to be tested for COVID-19, asymptomatic disease, and unreported at-home testing [3,14,15]. As a partner to traditional public health responses, WBE has shown to be a valuable tool in predicting and assaying case counts across populations both small and large [4,7,17,18]. Though WBE has shown to be a vital component of the world's fight against COVID-19, its major method of RT-qPCR on extracted RNA from wastewater samples is only able to quantify viral loads and cannot monitor the evolution of SARS-CoV-2 and the resulting viral variants.

The SARS-CoV-2 virus has mutated many times since its original genomic description, representing ongoing evolution as the COVID-19 pandemic progresses [19–22]. The WHO and PANGO Network monitor these mutations, broadly classifying variants into variants of concern (i.e., Alpha, Beta, Gamma, Delta, Omicron), variants of interest (i.e., Epsilon, Zeta, Eta, Mu), or variants under monitoring, along with lineage designations based on phylogenetics (i.e., B.1.351, B.1.1.529) [21–24]. In many cases, SARS-CoV-2 variants may possess mutations that confer phenotypic changes in the COVID-19 disease, such as increased transmissibility or antibody escape [19,25]. While these variants often contain numerous mutations, single nucleotide polymorphisms (SNPs) have been shown to occur across the SARS-CoV-2 genome, (known as single nucleotide variants; SNVs) representing mutational events that often have unknown functional or evolutionary consequences [18,26–29]. Currently, direct sampling from COVID-19-positive patients is the gold standard for SARS-CoV-2 sequencing but remains limited by the logistics required to administer tests and only allows for the sequencing of the virus from one patient at a time. Likewise, single-isolate patient sequencing likely misses rare SARS-CoV-2 variants [30], or those infecting non-human hosts, which may serve as undetected reservoirs for SNVs [31]. By sequencing SARS-CoV-2 from wastewater, we can capture circulating variants/SNVs across wide areas, which provide a composite sample representing large populations and may detect SNVs before standard medical sampling [18,26,30,32].

There are many challenges to sequencing SARS-CoV-2 from wastewater samples [33–35]. As a matrix of industrial, agricultural, and human-borne wastes, wastewater often contains a variety of detergents and other compounds that serve as PCR inhibitors and likely degrade viral particles [30,33,36,37]. Similarly, SARS-CoV-2 is often at a low viral load, and the virus detected in wastewater is almost certainly fragmented, making sequencing difficult due to an inability to cover an entire genome in one assay [18,38–40]. Sequencing methods have been developed to address these challenges, such as viral enrichment, targeted amplification of viral regions, and various RNA extraction protocols but many of these methods are designed for clinical samples, so wastewater analyses remain technically difficult to accurately conduct [17,18,41,42]. In order to increase our confidence in SARS-CoV-2 variant analyses, we used two sequencing library preparation methods on wastewater samples: The Illumina Respiratory Virus Oligonucleotide Panel, which enriches for respiratory virus nucleic acids before sequencing, and the QIAseq SARS-CoV-2 Primer Panel, which uses 200 PCR primer sets to amplify the entire SARS-CoV-2 genome.

Sequencing SARS-CoV-2 obtained from wastewater is a critical component of monitoring the ongoing COVID-19 pandemic [4]. Here, we present a study in which we used metatranscriptomic sequencing and two methods of library preparation (Illumina Respiratory Virus Oligonucleotide Panel or QIAseq SARS-CoV-2 Primer Panel) to identify SNVs, clades, and sublineages of SARS-CoV-2 on 317 influent wastewater samples. These samples were obtained from eight WTPs across Southern California from April 2020 to August 2021 and represent the collective wastewater of approximately 16 million residents. We investigated several lines of inquiry through our study: First, what RNA viruses are represented in our samples? Second, what clades and sublineages of SARS-CoV-2 were present in Southern California wastewater, and can we detect variants of concern in wastewater? Third, what SNVs were present in Southern California wastewater, and can we detect these variants with both library preparation methods? Lastly, does wastewater sequencing allow for the early detection of variants before clinical sequencing?

2. Materials and Methods

2.1. Sample Collection and Handling

We previously reported the sample collection and handling procedure in Rothman et al., 2021 [18], and Rothman et al., 2022 [43]. Briefly, we collected 317 1-L 24-h composite influent wastewater samples by autosampler at eight WTPs across Southern California between April 2020 and August 2021 (Table 1). We aliquoted and stored 50 mL of sample at 4 °C until processing.

Table 1. Sample quantities, date spans of collection, and approximate influent flow and served population. WTP names include the abbreviations used throughout the study, and “IRV” denotes Illumina Respiratory Virus Enrichment library preparation.

Wastewater Treatment Plant	Library Preparation Method	Number of Samples	Date Span	Approximate Inflow (Million Gallons/Day)	Approximate Population Served
Escondido Hale Avenue Resource Recovery Facility (ESC)	IRV	47	20 July 2020–19 July 2021	14	190,000
Hyperion Water Reclamation Plant (HTP)	IRV	89	11 August 2020–29 July 2021	275	4,000,000
	Tiled amplicon	47	21 April 2020–25 July 2021		
Joint Water Pollution Control Plant (JWPCP)	IRV	15	11 August 2020–17 November 2020	400	4,800,000
North City Water Reclamation Plant (NC)	IRV	5	14 August 2020–23 October 2020	30	1,400,000
Orange County Reclamation Plant #1 (OC)	IRV	14	12 August 2020–21 December 2020	140	2,600,000
Point Loma Water Treatment Plant (PL)	IRV	77	13 August 2020–3 August 2021	175	2,200,000
	Tiled amplicon	48	30 April 2020–18 August 2021		
San Jose Creek Water Reclamation Plant (SJ)	IRV	15	12 August 2020–18 November 2020	100	1,000,000
South Bay Water Reclamation Plant (SB)	IRV	7	13 August 2020–5 November 2020	15	107,000

2.2. Wastewater Sample RNA Extraction

We used two separate RNA extraction and library preparation protocols for the samples in this study. For one set (N = 269), we used a protocol based on Crits-Christoph, 2021 [26], and Rothman, 2021 [18], and we refer to those papers for detailed RNA extraction methods. Briefly, 50 mL of influent wastewater was pasteurized at 65 °C for 90 min in a water bath, filtered through 0.22- μ M sterile filters (VWR, Radnor, PA, USA), then centrifugated at 3000 \times g with 10 kDa filters (MilliporeSigma, Burlington, MA, USA) and RNA was extracted with an Invitrogen PureLink RNA Mini Kit plus DNase (Invitrogen, Waltham, MA, USA). We will refer to these samples as “Illumina Respiratory Virus-enriched” (IRV) henceforth.

A second set of RNA extractions was carried out (N = 95) with a different extraction and library preparation protocol based on Steele, 2021 [35]. Briefly, we added 25 mM MgCl₂ to 20 mL of wastewater, then acidified the samples to pH < 3.5 with HCl. We then

transferred the mixture to a cellulose ester membrane (type HA; Millipore, Bedford, MA, USA) then bead bashed the filters in preloaded 2 mL ZR BashingBead lysis tubes (Zymo, Irvine, CA, USA) for 1 min. Lastly, we extracted total nucleic acid with a NucliSENS extraction kit with magnetic bead capture following the supplied protocol (bioMérieux, Durham, NC, USA). Libraries for these samples were then prepared as follows and will be referred to as “tiled amplicons”. Note that some samples were only sequenced with the IRV method, while others were only sequenced with the tiled amplicon method. A total of $N = 47$ samples ($N = 18$ Hyperion and $N = 29$ Point Loma) were sequenced with both methodologies.

2.3. Sequencing Library Preparation

All sample library preparation and sequencing steps were carried out by the University of California Irvine Genomics High-Throughput Facility (GHTF). The GHTF prepared IRV-enriched libraries with the Illumina Respiratory Virus Oligonucleotide panel paired with an Illumina RNA prep with enrichment kit (Illumina, San Diego, CA, USA), following the manufacturer’s protocol. The tiled amplicon libraries were prepared by using the QIAseq SARS-CoV-2 Primer Panel paired with a QIAseq FX DNA Library UDI kit (Qiagen, Germantown, MD, USA) and using the manufacturer’s protocol. The GHTF sequenced the resulting paired-end libraries as either 2×100 bp or 2×150 bp (Supplemental File SF1) on an Illumina NovaSeq 6000 with an S4 300 cycling kit and sent the data as demultiplexed FASTQ files.

2.4. Bioinformatics and Sequence Data Processing

All data processing was conducted on the UCI High-Performance Community Computing Cluster (HPC3). We removed sequencing adapter sequences and low-quality bases with the BBTools software v.38.87 “bbduk” [44]. We subsequently marked sequencing duplicates with Picard toolkit “MarkDuplicates” [45], removed reads mapping to the HG38 human genome with Bowtie2 [46], then used Kraken2 [47] and Bracken [48] to taxonomically classify our reads for reporting purposes and plotted those relative abundances as a stacked bar plot with the R package “ggplot2” [49,50].

Once we had removed the human reads, we aligned the reads to the SARS-CoV-2 Hu-1 reference strain [51] with Bowtie2, then sorted and indexed the resulting bam files with “samtools” [52]. We used iVar [53] with the default settings to trim off QIAseq primer sequences and call single nucleotide variants (SNVs) with a Fisher’s exact test of $p < 0.05$, as compared to the reference strain (Supplemental File SF1). Subsequently, we used Freyja [30] to assign SARS-CoV-2 lineage and sublineage identities to the alignments using the USHER phylogeny [54] and then de-mix the clades and sublineages within each sample to calculate approximate relative abundances. As we wanted to compare the results from IRV-enriched libraries to tiled amplicon libraries, we also used the iVar/Freyja pipeline to call SNVs and assign lineages/sublineages to these libraries, even though there were no true primers to remove.

To compare our wastewater sequence data to clinical sequencing, we obtained the date of sublineage detection and reported genomes from PANGO, GISAID, and the California Health and Human Services Agency [21,22,55,56] and associated these dates with our wastewater sampling dates. Due to the longitudinal nature of our data, we compared variant lineage abundances (as counts per million) over time with MaAsLin2 [57] where we had yearlong data, using WTP and sequencing batch as random effects. Likewise, we reported SNVs from 68 samples with detectable SARS-CoV-2 in Rothman 2021 [18] but here we reanalyzed the data and present the new results for consistency with our new methods.

We investigated SARS-CoV-2 sublineage alpha diversity through Kruskal–Wallis testing and beta diversity through Adonis PERMANOVA testing with the R package “vegan” [58] and measured the change in diversity with linear mixed-effects regression models (LMERs) with the R package “lmerTest” [59] using both WTP and sequencing batch as random effects. Lastly, we plotted all of the data with the R packages “ggplot2”, “ggre-

pel” [60], “Rcartocolor” [61], and “Patchwork” [62], and created the graphical abstract with BioRender.com.

2.5. Data Availability

Representative analyses’ scripts and code are available at https://github.com/jasonarothman/wastewater_sarscov2_apr20_aug21, and raw sequencing files are deposited at the NCBI Sequence Read Archive under accession number PRJNA729801. SARS-CoV-2 lineage assignments and SNV calls are available in Supplemental File SF1, the California Health and Human Service Agency COVID-19 Variants Dataset [56], and GISAID data are available by request from GISAID (<https://gisaid.org/>) [55] per their terms of use.

3. Results

3.1. Library Statistics, Sample Composition, and SARS-CoV-2 Genome Coverage

We used two library preparation techniques on our samples (IRV-enriched and tiled amplicon), so we report the summary statistics separately below. For IRV-enriched samples, we sequenced 548,883,572 non-human quality-filtered paired-end reads (average = 1,020,230, range = 9910–8,243,363) across 269 samples. We taxonomically classified an average of 58.8% of reads (range = 8.9–85.4%), of which an average of 9.4% of overall reads were viral (range = 0.1–53.7%). Of total viruses, 2,281,212 reads (6.0%) mapped to SARS-CoV-2. Regarding tiled-amplicon samples, we sequenced 1,074,798,497 non-human quality-filtered paired-end reads (average = 11,313,668, range = 8,619,210–14,384,197) across 95 samples. We classified an average of 47.2% of reads (range = 35.9–66.7%), of which an average of 4.5% were viral (range ≤ 0.01–41.6%). Of the total tiled-amplicon prepared viruses, 47,427,550 reads (99.6%) mapped to SARS-CoV-2 (Figure 1).

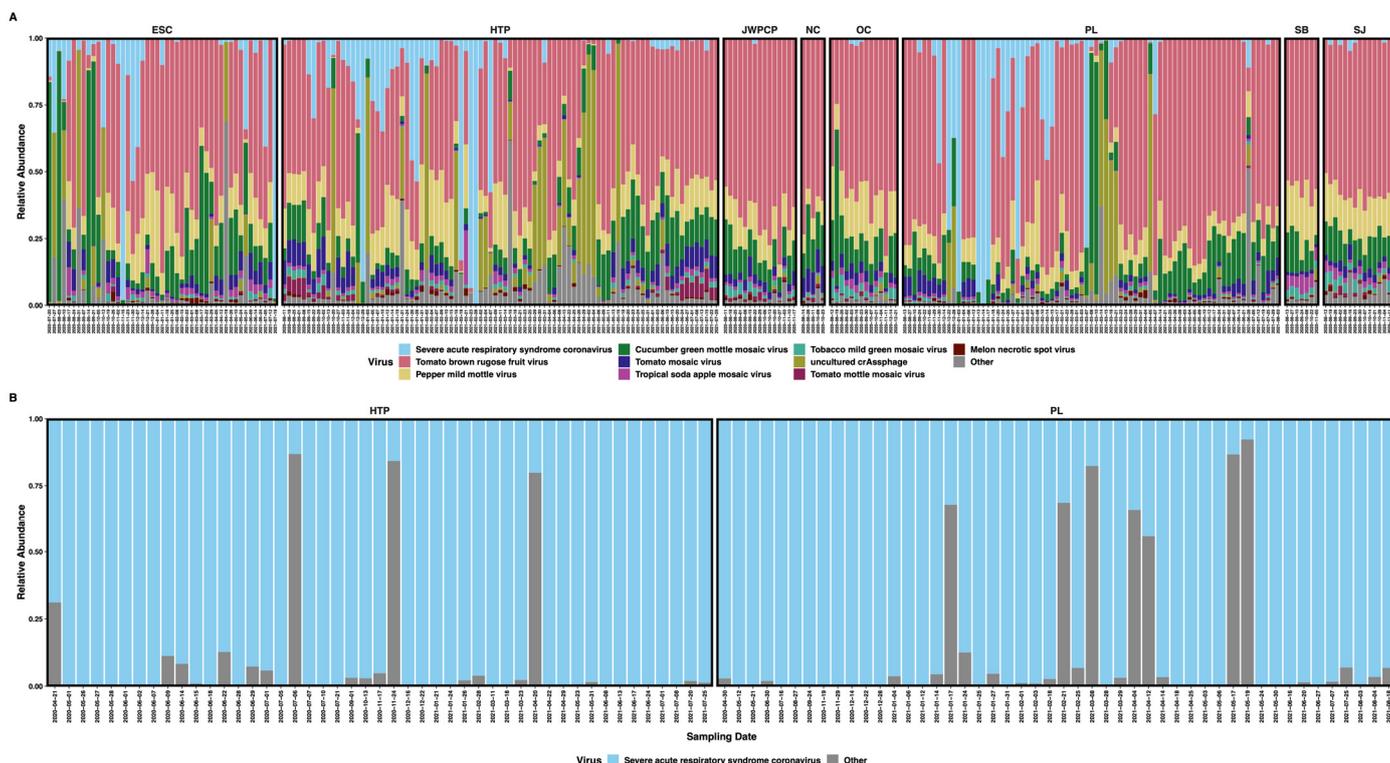


Figure 1. Stacked bar plots showing the relative abundances of RNA reads mapping to (A) the top 10 most proportionally abundant viruses plus all others in respiratory virus-enriched libraries and (B) SARS-CoV-2 plus other viruses in tiled-amplicon libraries. Plots are faceted by WTP and labeled with sampling date.

We obtained broad SARS-CoV-2 genome coverage with both sequencing approaches. When considering all samples together, IRV-prepared libraries covered 99.92% of the SARS-CoV-2 genome at a mean sequencing depth of only $2\times$ at each base position. Tiled amplicon library preparation had both wider coverage and higher sequencing depth, with these libraries covering 99.95% of the genome at a mean depth of 76 reads per base (Figure S1).

3.2. SARS-CoV-2 Read Classification

Because SARS-CoV-2 was well-represented in our samples, we could classify reads mapping to SARS-CoV-2 through UShER SARS-CoV-2 barcoding and de-mixing to approximate relative abundances with Freyja. We were able to classify many of the mapped reads to specific named variants of interest (VOIs) and variants of concern (VOCs) (both currently circulating and historically significant) along with other sublineages of SARS-CoV-2 that do not correspond to a VOI/VOC. Within tiled-amplicon samples that had classifiable reads ($N = 90$), these “top 10” most proportionally abundant clades were Alpha ($\bar{x} = 3.3\%$, range = 0–98.8%), Beta ($\bar{x} = 0.23\%$, range = 0–17.2%), Delta ($\bar{x} = 16.0\%$, range 0–100%), Epsilon ($\bar{x} = 4.6\%$, range = 0–98.9%), Gamma ($\bar{x} = 0.02\%$, range = 0–14.3%), Iota ($\bar{x} = 0.02\%$, range = 0–0.7%), Lambda ($\bar{x} = 1.1\%$, range = 0–96.8%), Mu ($\bar{x} = 0.1\%$, range = 0–8.6%), Zeta ($\bar{x} = 0.01\%$, range = 0–0.3%), and all other sublineages combined ($\bar{x} = 67.7\%$, range = 0–99.9%) (Figure 2 and Figure S2). While individual samples contained varying proportions of WHO clades, only the relative abundance of the Delta variant was shown to increase throughout the course of the experiment ($\beta = 2.57$, $P_{\text{adj}} < 0.001$)

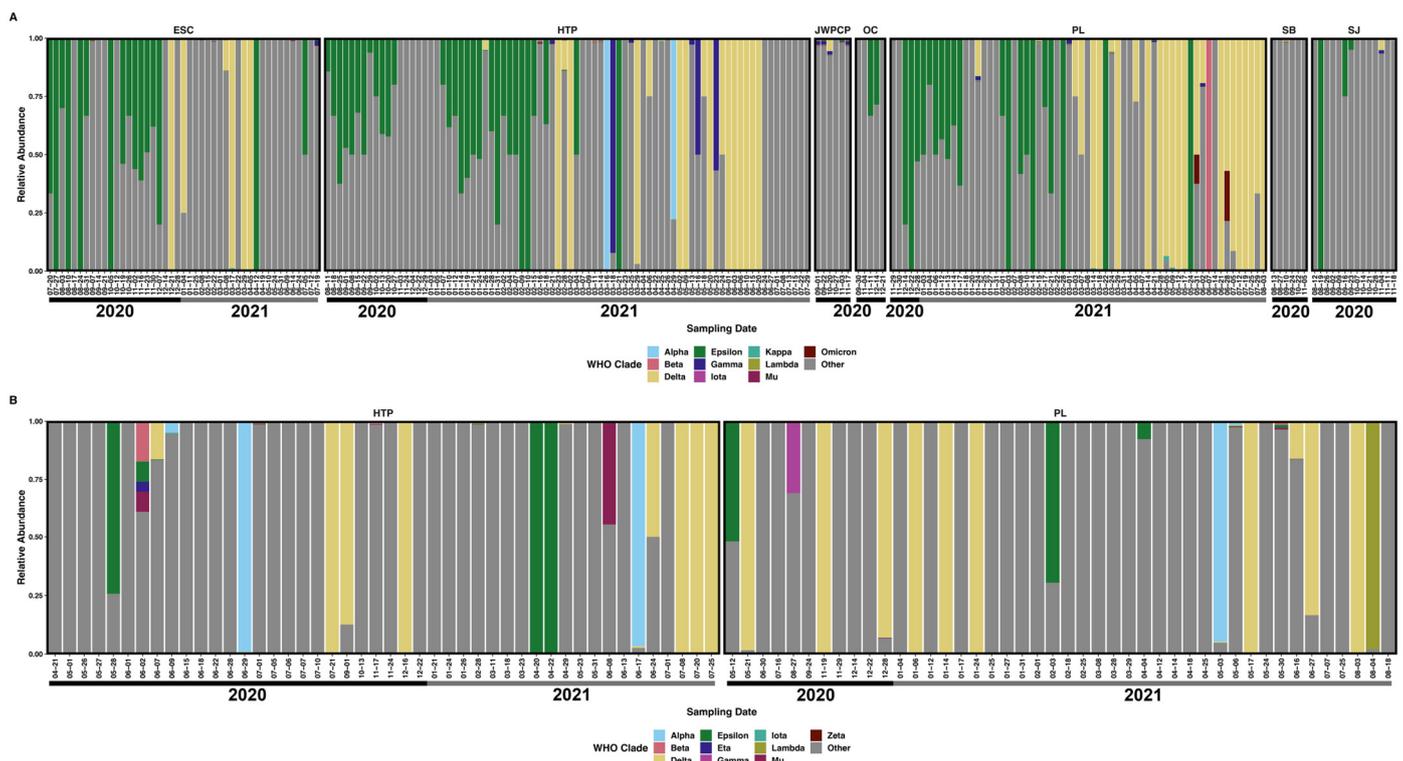


Figure 2. The relative proportional abundance of the ten most-abundant SARS-CoV-2 lineages plus others in (A) respiratory virus-enriched libraries, and (B) tiled-amplicon libraries faceted by WTP and labeled with sampling date. Note that one sample date from the North City Water Reclamation Plant is not shown.

Within IRV-enriched samples that had classifiable reads ($N = 219$), the “top 10” most proportionally abundant SARS-CoV-2 clades were largely similar to tiled-amplicon samples. These clades were Alpha ($\bar{x} = 0.86\%$, range = 0–100%), Beta ($\bar{x} = 0.50\%$, range = 0–100%),

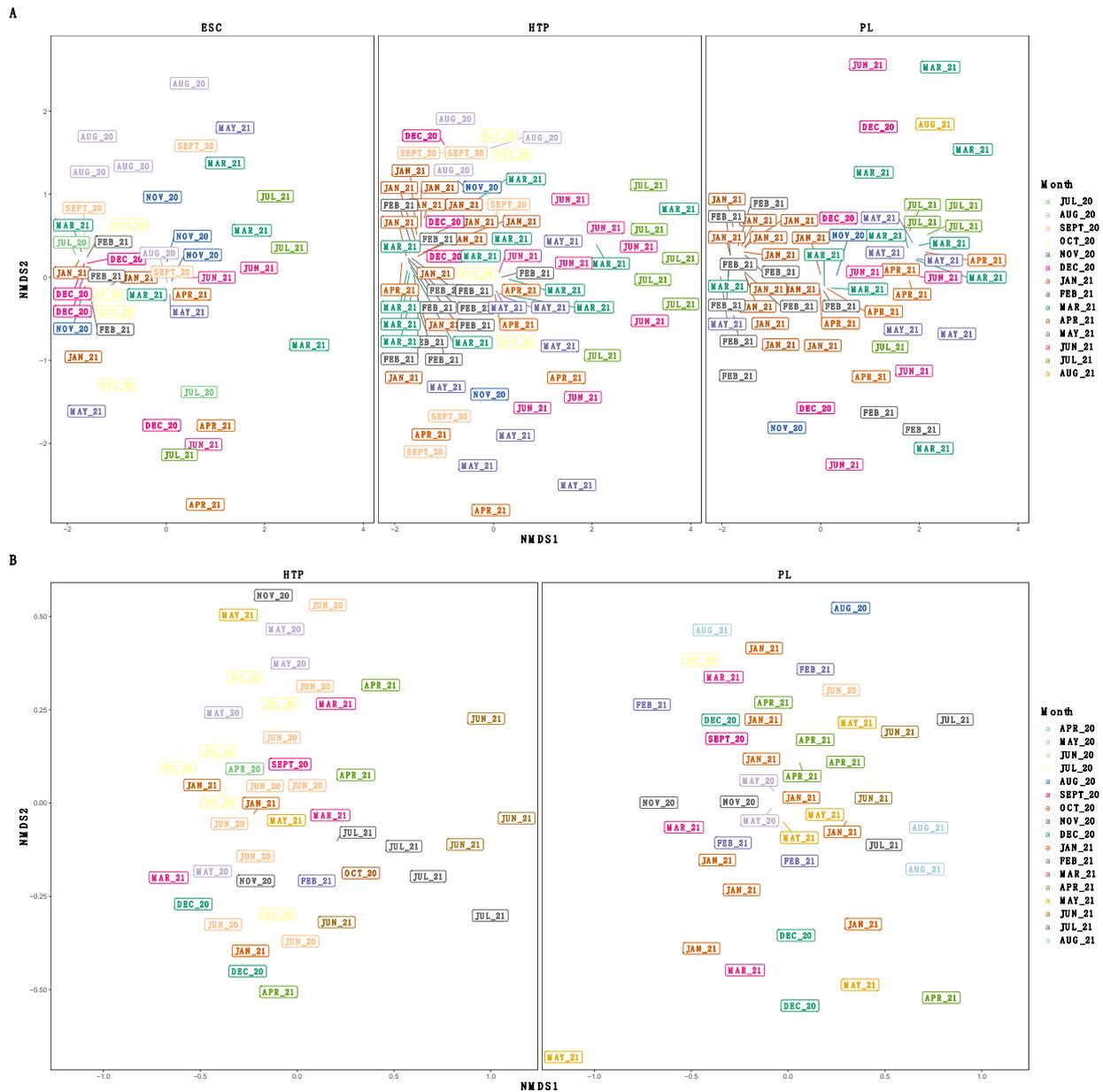


Figure 4. Non-metric multidimensional scaling (NMDS) ordinations of the Bray–Curtis dissimilarities of SARS-CoV-2 sublineages faceted by water treatment plant for (A) respiratory virus-enriched (IRV) and (B) tiled-amplicon libraries. SARS-CoV-2 sublineages did not significantly differ between WTPs (PERMANOVA [IRV: $p = 0.07$, $R^2 = 0.01$], [tiled-amplicon: $p = 0.58$, $R^2 = 0.01$]) but differed by calendar month (PERMANOVA [IRV: $p < 0.001$, $R^2 = 0.15$], [tiled-amplicon: $p < 0.001$, $R^2 = 0.21$]). Color and plot labels denote sampling month, and only WTPs with yearlong data are included.

We analyzed the tiled-amplicon samples in the same fashion as above, and did not find a difference in sublineage alpha diversity between WTP ($H_{(1)} = 0.04$, $p = 0.84$) or calendar month ($H_{(16)} = 11.2$, $p = 0.80$), and diversity remained constant over numerical time ($t = 0.41$, $p = 0.68$). We observed a difference in beta diversity by month ($R^2 = 0.21$, $p < 0.001$) but not between WTPs ($R^2 = 0.01$, $p = 0.58$), sequencing batches ($R^2 = 0.01$, $p = 0.20$), or an interaction between WTP and month ($R^2 = 0.13$, $p = 0.19$), nor by numerical time ($t = -1.8$, $p = 0.07$) (Figure 4).

As we were unable to collect samples from all eight WTPs for the full year, we also analyzed subclade diversity during months with the broadest WTP coverage (August–November 2020 without WTP “NC” as there was only one sample, $N = 61$). There was no difference in

sublineage alpha diversity between WTPs ($H_{(6)} = 11.9, p = 0.07$), months ($H_{(3)} = 2.5, p = 0.47$), or numerical time ($t = -1.4, p = 0.17$). There was a significant difference in sublineage beta diversity between WTPs ($R^2 = 0.15, p < 0.001$), and an interaction between WTP and month ($R^2 = 0.21, p < 0.027$), with no differences by calendar month ($R^2 = 0.06, p = 0.09$), sequencing batch ($R^2 = 0.01, p = 0.62$), or numerical time ($t = -0.8, p = 0.41$).

3.4. SARS-CoV-2 Single Nucleotide Variants

At a more granular level, we also tabulated SARS-CoV-2 single nucleotide variants (SNVs) throughout our samples (Supplemental File SF1). Combined, IRV-enriched and tiled-amplicon sample preparation methods detected 2871 SARS-CoV-2 SNVs across the genome, with each approach capturing a different number of SNVs, and in many cases, different genomic locations (Figure 5). IRV-enriched samples contained 1212 SNVs, most being found only once (1071) or twice (83); however, we often detected the same SNV multiple times at several genomic positions in separate samples (Figure 5). For example, SNVs at nucleotide positions 23403, 241, 14408, 17014, 3037, 28272, 8947, 12878, 2597, 21600, 25563, and 28887 were each detected over 10 times across the samples. Tiled-amplicon sequencing also detected SNVs well, identifying 1808 SNVs across the SARS-CoV-2 genome. Similar to IRV-enriched results, most SNVs were found once (1030) or twice (254), although several SNVs were identified in multiple samples (Figure 5): SNVs located at nucleotide positions 22796, 28971, 22656, 28982, and 9864 were detected over 40 times across tiled-amplicon samples.

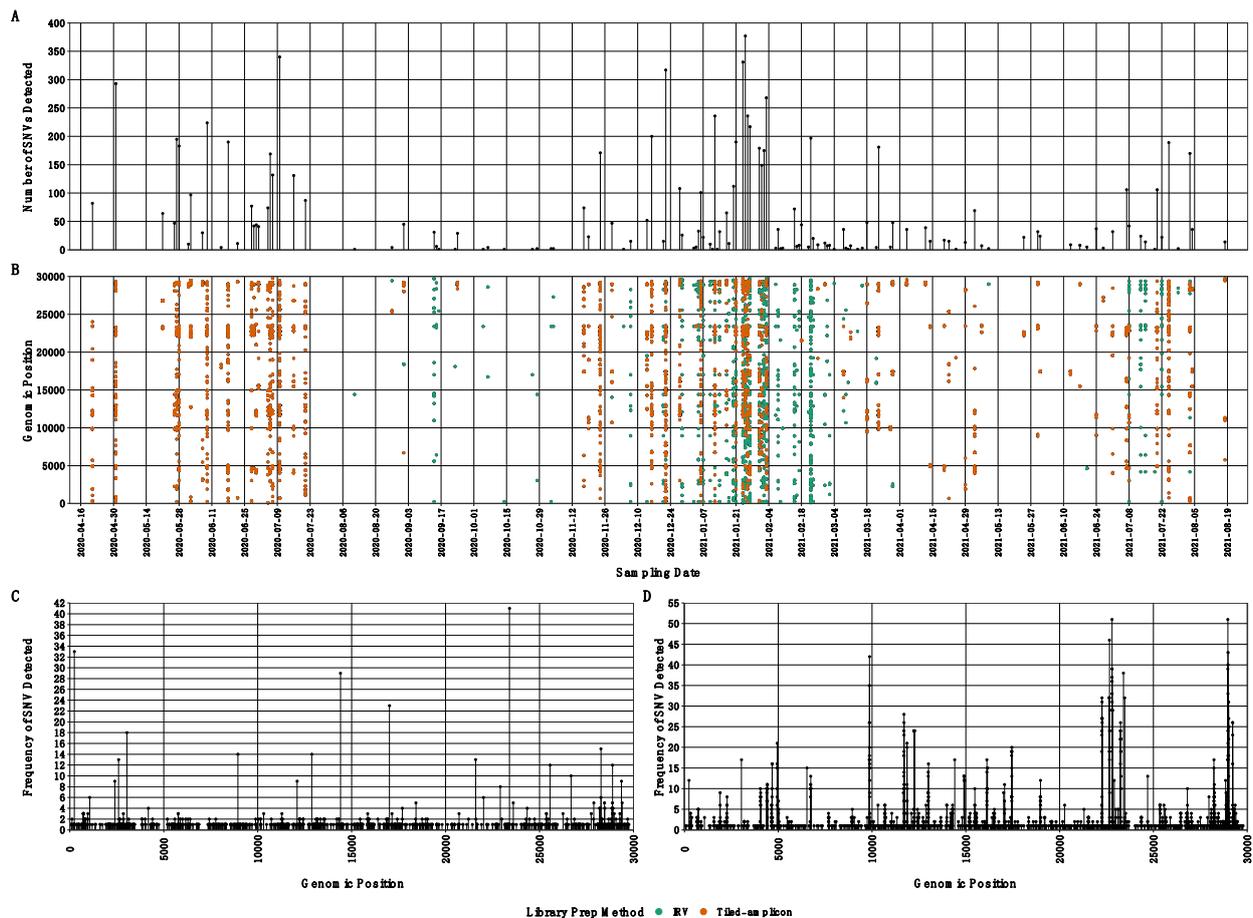


Figure 5. (A) Number of single nucleotide variants (SNVs) detected at each sample date and (B) nucleotide position across the SARS-CoV-2 genome for all samples colored by library preparation

method (IRV signifies Illumina Respiratory Virus enrichment panel); (C,D) indicate the frequency of SNVs detected at each position of the SARS-CoV-2 genome across all respiratory virus-enriched and tiled-amplicon libraries, respectively.

4. Discussion

Our study represents a large-scale effort to employ wastewater-based epidemiology (WBE) across a catchment area of 16 million people and supports the monitoring of SARS-CoV-2 evolution throughout the ongoing COVID-19 pandemic. Through respiratory virus-enriched and tiled-amplicon RNA sequencing approaches, we classified SARS-CoV-2 lineages and single nucleotide variants (SNVs) and could approximate VOCs/VOIs across a yearlong study of Southern California wastewater. Like other studies, we captured SARS-CoV-2 mutations across the genome and show the potential to detect sublineages and SNVs months before clinical analyses of patient samples [18,26,27,30,39]. While WBE is a powerful tool—and is not subject to many of clinical sequencing’s drawbacks—we cannot use these methods to determine the exact source of SARS-CoV-2 variants [18,26,30,38] and instead propose the use of WBE to monitor populations instead of individuals. Our results suggest that multi-scale sampling of individual patients, local wastewater catchments (i.e., university campuses), and WTPs can give public health agencies vital information to identify novel SARS-CoV-2 variants and predict disease spread to further combat COVID-19 [4,11,30].

4.1. Classifying SARS-CoV-2 Reads and Comparing Wastewater to Clinical Sequencing

In most samples, we could classify SARS-CoV-2 RNA fragments at multiple levels of resolution—both at the named variants (i.e., Alpha, Beta, etc.) and sublineage levels (i.e., B.1.429, B.1.617.2, P.1, etc.)—and calculate semi-quantitative relative abundances. When considering the full year of data, sublineage diversity of SARS-CoV-2 was not different between WTPs, rather it changed monthly probably due to the similarity of proportional disease burden and proximity of the San Diego and Los Angeles counties [1]. As expected, our sublineage quantification was not exactly concordant with clinical sequencing data, probably due to the aggregate nature of wastewater and our composite sampling, along with the lack of clinical specimens early in the pandemic [4,21,22,30,38]. We do note, however, that ours and clinical data agree well during the emergence of the Delta variant, suggesting that wastewater can detect the potential evolutionary replacement of lineages accurately as has been recently shown with the domination of the Omicron variant [30,63]. Additionally, when comparing our sequencing data specifically to California clinical data, we observed a higher apparent concordance between named SARS-CoV-2 clades, indicating that WBE is powerful on smaller scales, which may assist public health agencies with more targeted responses to disease. Similarly, we detected many SARS-CoV-2 sublineages earlier in wastewater than clinical sequencing—in some cases by several months—further supporting work indicating that WBE is useful for predicting disease load and the spread of novel variants [30,63]. Naturally, we recognize this is a retrospective study, and rely on clinical sequencing to name and prioritize the variants we sequenced in wastewater, so we suggest that public health and wastewater sequencing be used in tandem to carefully monitor the evolutionary potential of SARS-CoV-2 [4].

4.2. Identifying SARS-CoV-2 Single Nucleotide Variants

Similar to previous work, we detected thousands of single nucleotide variants (SNVs) across samples and sequenced putatively novel or rare SNVs that have an unknown function or species host [18,26,27,31,63,64]. For example, in many samples (from within or between WTPs), we detected SNVs at positions 9864, 22796, and 28971, which are exceedingly rare in public sequencing data, along with SNVs at 241, 14408, and 23403, which were common in 2020 [65,66]. Our ability to detect both low-prevalence and near-ubiquitous SNVs indicates that WBE is broadly useful for accurate SNV detection and may provide a reasonable estimate of what SNVs are circulating across populations [26,30,31,63,64].

Likewise, when comparing our results to other wastewater studies, we detected variants or sublineages also reported in Nice, New York, Montana, Arizona, Northern California, Berlin, and across Austria, often at similar sampling dates, which shows that sequencing wastewater is reproducible and accurate at very large scales [26,31,39,67,68]. Being that sequencing wastewater is technically challenging, we qualitatively compared two major methods of SARS-CoV-2 analysis and note that targeted amplification provided better sequencing depth and resolution, indicating its utility when presented with degraded low-titer RNA and the detergent/PCR inhibitor content of wastewater along with our harsh extraction methods [4,18,41]. Nevertheless, there is the possibility that targeted amplification may miss novel variants due to mutational differences in SARS-CoV-2 genomes [69,70], something suggested by our IRV method detecting Omicron where tiled amplicon did not. We also recognize that extracted RNA with two different methods, which likely confounds our results but we still suggest the use of targeted amplification approaches for wastewater samples, which supports previous work and method development [4,18,26,30,41,63,71,72].

5. Conclusions

Wastewater-based epidemiology has exploded into a worldwide endeavor and is a critical part of humanity's response to the COVID-19 pandemic [4,6,9,11]. Our study demonstrates WBE's effectiveness in monitoring SARS-CoV-2 mutations across megaregions [18,26,30,31,39], which continues to be important as novel VOCs emerge and the popularity of at-home testing reduces public health's ability to accurately quantify COVID-19 cases [3,14,15]. As WBE technology matures, future directions for method development should address the need to simultaneously detect and sequence variants from other circulating diseases—such as influenza, Respiratory Syncytial Virus, and Norovirus—alongside SARS-CoV-2. COVID-19 has demonstrated the need for scientists, wastewater agencies, and public health to work together to track the evolution and spread of SARS-CoV-2, especially in underserved areas, low population coverage, and places where the medical field is overburdened [4,16]. WBE has the potential to discover emergent diseases and should be implemented across population centers as a sentinel for the next pandemic.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/applmicrobiol4020044/s1>, Figure S1: Area plot of the average reads-per-base mapped to the SARS-CoV-2 genome across all samples. Panel (A) represents libraries prepared with the Illumina Respiratory Virus Enrichment Panel (IRV) and Panel (B) represents libraries prepared for tiled amplicon sequencing. Figure S2: The relative proportional abundance of the ten most abundant SARS-CoV-2 lineages plus others in (A) tiled-amplicon libraries, (B) weekly GISAID-reported data for clinical samples from the United States, and (C) California Health and Human Services (CalHHS)-reported clinical data. Panel (A) is faceted by WTP and labeled with sampling date and panel B is labeled by the aggregate of GISAID-reported data. Note that one sample date from the North City Water Reclamation Plant is not shown, and variant data from CalHHS are not available before 1 January 2021 so those dates are left blank. Figure S3: The relative proportional abundance of the ten most abundant SARS-CoV-2 lineages plus others in Illumina Respiratory Virus Panel-enriched libraries for (A) Escondido, (B) Hyperion, and (C) Point Loma water treatment facilities. Panels (D–F) are weekly GISAID-reported data for clinical samples from the United States, corresponding to the approximate time periods of panels (A–C), respectively. Panels (G–I) are CalHHS-reported clinical data corresponding to the approximate time periods of panels (A–C), respectively. Variant data from CHSS are not available before 1 January 2021 so those dates are left blank. Supplemental File SF1: This file contains several worksheets of information: A README sheet with detailed information, sample metadata, single nucleotide variants for IRV and tiled amplicon samples, and PANGO sublineages as output by Freyja for IRV and tiled amplicon samples.

Author Contributions: Conceptualization, J.A.R., A.G.Z.-F., J.A.S., J.F.G. and K.L.W.; methodology, J.A.R., A.G.Z.-F., J.A.S., J.F.G. and K.L.W.; validation, J.A.R.; formal analysis, J.A.R. and K.L.; investigation, J.A.R., A.S., K.L. and K.R.; resources, J.A.R., A.G.Z.-F., J.A.S., J.F.G., K.L. and K.R.; data curation, J.A.R. and K.L.; writing—original draft preparation, J.A.R. and K.L.W.; writing—review and editing,

J.A.R., A.G.Z.-F., J.A.S., J.F.G. and K.L.W.; visualization, J.A.R.; supervision, J.A.R., A.G.Z.-F., J.A.S., J.F.G. and K.L.W.; project administration, J.A.R., J.F.G. and K.L.W.; funding acquisition, J.A.R. and K.L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the University of California Office of the President Research Grants Program Office (award numbers R01RG3732 and R00RG2814) awarded to JAR and KLW, and a Hewitt Foundation for Biomedical Research postdoctoral fellowship to JAR. This work was made possible, in part, through access to the Genomics High-Throughput Facility Shared Resource of the Cancer Center Support Grant (P30CA-062203) at the University of California, Irvine, NIH shared instrumentation grants 1S10RR025496-01, 1S10OD010794-01, and 1S10OD021718-01, and access to computing resources from the UCI High-Performance Community Computing Cluster.

Data Availability Statement: Representative analyses' scripts and code are available at https://github.com/jasonarothman/wastewater_sarscov2_apr20_aug21, and raw sequencing files are deposited at the NCBI Sequence Read Archive under accession number PRJNA729801. SARS-CoV-2 lineage assignments and SNV calls are available in Supplemental File SF1, the California Health and Human Service Agency COVID-19 Variants Dataset [56], and GISAID data are available by request from GISAID (<https://gisaid.org/>) [55] per their terms of use.

Acknowledgments: We thank the Los Angeles and Orange County Sanitation Districts, the City of San Diego Public Utilities, the City of Escondido Hale Avenue Resource Recovery Facility, and the City of Los Angeles Department of Sanitation and Environment for collecting wastewater samples. We also thank the developers of Freyja and Adélaïde Roguet for software assistance, and Seung-Ah Chung for library preparation assistance.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Dong, E.; Du, H.; Gardner, L. An Interactive Web-Based Dashboard to Track COVID-19 in Real Time. *Lancet Infect. Dis.* **2020**, *20*, 533–534. [[CrossRef](#)] [[PubMed](#)]
- Chan, J.F.-W.; Kok, K.-H.; Zhu, Z.; Chu, H.; To, K.K.-W.; Yuan, S.; Yuen, K.-Y. Genomic Characterization of the 2019 Novel Human-Pathogenic Coronavirus Isolated from a Patient with Atypical Pneumonia after Visiting Wuhan. *Emerg. Microbes Infect.* **2020**, *9*, 221–236. [[CrossRef](#)] [[PubMed](#)]
- Reiner, R.C.; Barber, R.M.; Collins, J.K.; Zheng, P.; Adolph, C.; Albright, J.; Antony, C.M.; Aravkin, A.Y.; Bachmeier, S.D.; Bang-Jensen, B.; et al. Modeling COVID-19 Scenarios for the United States. *Nat. Med.* **2021**, *27*, 94–105.
- Wu, F.; Lee, W.L.; Chen, H.; Gu, X.; Chandra, F.; Armas, F.; Xiao, A.; Leifels, M.; Rhode, S.F.; Wuertz, S.; et al. Making Waves: Wastewater Surveillance of SARS-CoV-2 in an Endemic Future. *Water Res.* **2022**, *219*, 118535. [[CrossRef](#)]
- Leite, H.; Lindsay, C.; Kumar, M. COVID-19 Outbreak: Implications on Healthcare Operations. *TQM J.* **2020**, *39*, 88. [[CrossRef](#)]
- Sharara, N.; Endo, N.; Duvallet, C.; Ghaeli, N.; Matus, M.; Heussner, J.; Olesen, S.W.; Alm, E.J.; Chai, P.R.; Erickson, T.B. Wastewater Network Infrastructure in Public Health: Applications and Learnings from the COVID-19 Pandemic. *PLoS Glob. Public Health* **2021**, *1*, e0000061. [[CrossRef](#)]
- Karthikeyan, S.; Ronquillo, N.; Belda-Ferre, P.; Alvarado, D.; Javidi, T.; Longhurst, C.A.; Knight, R. High-Throughput Wastewater SARS-CoV-2 Detection Enables Forecasting of Community Infection Dynamics in San Diego County. *mSystems* **2021**, *6*, e00045-21. [[CrossRef](#)]
- Singer, A.C.; Thompson, J.R.; Filho, C.R.M.; Street, R.; Li, X.; Castiglioni, S.; Thomas, K.V. A World of Wastewater-Based Epidemiology. *Nat. Water* **2023**, *1*, 408–415. [[CrossRef](#)]
- Naughton, C.C.; Roman, F.A., Jr.; Alvarado, A.G.F.; Tariqi, A.Q.; Deeming, M.A.; Bibby, K.; Bivins, A.; Rose, J.B.; Medema, G.; Ahmed, W.; et al. Show Us the Data: Global COVID-19 Wastewater Monitoring Efforts, Equity, and Gaps. *FEMS Microbes* **2023**, *4*, xtad003. [[CrossRef](#)]
- Peccia, J.; Zulli, A.; Brackney, D.E.; Grubaugh, N.D.; Kaplan, E.H.; Casanovas-Massana, A.; Ko, A.I.; Malik, A.A.; Wang, D.; Wang, M.; et al. Measurement of SARS-CoV-2 RNA in Wastewater Tracks Community Infection Dynamics. *Nat. Biotechnol.* **2020**, *38*, 1164–1167. [[CrossRef](#)]
- Bivins, A.; North, D.; Ahmad, A.; Ahmed, W.; Alm, E.; Been, F.; Bhattacharya, P.; Bijlsma, L.; Boehm, A.B.; Brown, J.; et al. Wastewater-Based Epidemiology: Global Collaborative to Maximize Contributions in the Fight against COVID-19. *Environ. Sci. Technol.* **2020**, *54*, 7754–7757. [[CrossRef](#)] [[PubMed](#)]
- Jarrom, D.; Elston, L.; Washington, J.; Prettyjohns, M.; Cann, K.; Myles, S.; Groves, P. Effectiveness of Tests to Detect the Presence of SARS-CoV-2 Virus, and Antibodies to SARS-CoV-2, to Inform COVID-19 Diagnosis: A Rapid Systematic Review. *BMJ Evid. Based Med.* **2022**, *27*, 33–45. [[CrossRef](#)] [[PubMed](#)]
- Lu, X.; Wang, L.; Sakthivel, S.K.; Whitaker, B.; Murray, J.; Kamili, S.; Lynch, B.; Malapati, L.; Burke, S.A.; Harcourt, J.; et al. US CDC Real-Time Reverse Transcription PCR Panel for Detection of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerg. Infect. Dis.* **2020**, *26*, 1654. [[CrossRef](#)] [[PubMed](#)]

14. Whittaker, C.; Walker, P.G.T.; Alhaffar, M.; Hamlet, A.; Djaafara, B.A.; Ghani, A.; Ferguson, N.; Dahab, M.; Checchi, F.; Watson, O.J. Under-Reporting of Deaths Limits Our Understanding of True Burden of COVID-19. *BMJ* **2021**, *375*, n2239. [[CrossRef](#)] [[PubMed](#)]
15. Qasmieh, S.A.; Robertson, M.M.; Teasdale, C.A.; Kulkarni, S.G.; McNairy, M.; Borrell, L.N.; Nash, D. The Prevalence of SARS-CoV-2 Infection and Uptake of COVID-19 Antiviral Treatments during the BA.2/BA.2.12.1 Surge, New York City, April-May 2022. *Commun. Med.* **2023**, *3*, 92. [[CrossRef](#)] [[PubMed](#)]
16. Reitsma, M.B.; Claypool, A.L.; Vargo, J.; Shete, P.B.; McCorvie, R.; Wheeler, W.H.; Rocha, D.A.; Myers, J.F.; Murray, E.L.; Bregman, B.; et al. Racial/Ethnic Disparities In COVID-19 Exposure Risk, Testing, And Cases At The Subcounty Level In California. *Health Aff.* **2021**, *40*, 870–878. [[CrossRef](#)] [[PubMed](#)]
17. Wu, F.; Xiao, A.; Zhang, J.; Moniz, K.; Endo, N.; Armas, F.; Bonneau, R.; Brown, M.A.; Bushman, M.; Chai, P.R.; et al. SARS-CoV-2 RNA Concentrations in Wastewater Foreshadow Dynamics and Clinical Presentation of New COVID-19 Cases. *Sci. Total Environ.* **2022**, *805*, 150121. [[CrossRef](#)] [[PubMed](#)]
18. Rothman, J.A.; Loveless, T.B.; Kaptcia, J., 3rd; Adams, E.D.; Steele, J.A.; Zimmer-Faust, A.G.; Langlois, K.; Wanless, D.; Griffith, M.; Mao, L.; et al. RNA Viromics of Southern California Wastewater and Detection of SARS-CoV-2 Single-Nucleotide Variants. *Appl. Environ. Microbiol.* **2021**, *87*, e0144821. [[CrossRef](#)] [[PubMed](#)]
19. Tao, K.; Tzou, P.L.; Nouhin, J.; Gupta, R.K.; de Oliveira, T.; Kosakovsky Pond, S.L.; Fera, D.; Shafer, R.W. The Biological and Clinical Significance of Emerging SARS-CoV-2 Variants. *Nat. Rev. Genet.* **2021**, *22*, 757–773. [[CrossRef](#)]
20. Badua, C.L.D.C.; Baldo, K.A.T.; Medina, P.M.B. Genomic and Proteomic Mutation Landscapes of SARS-CoV-2. *J. Med. Virol.* **2021**, *93*, 1702–1721. [[CrossRef](#)]
21. O’Toole, Á.; Scher, E.; Underwood, A.; Jackson, B.; Hill, V.; McCrone, J.T.; Colquhoun, R.; Ruis, C.; Abu-Dahab, K.; Taylor, B.; et al. Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool. *Virus Evol.* **2021**, *7*, veab064. [[CrossRef](#)] [[PubMed](#)]
22. Rambaut, A.; Holmes, E.C.; O’Toole, Á.; Hill, V.; McCrone, J.T.; Ruis, C.; du Plessis, L.; Pybus, O.G. A Dynamic Nomenclature Proposal for SARS-CoV-2 Lineages to Assist Genomic Epidemiology. *Nat. Microbiol.* **2020**, *5*, 1403–1407. [[CrossRef](#)]
23. World Health Organization Tracking SARS-CoV-2 Variants. Available online: <https://www.who.int/activities/tracking-SARS-CoV-2-variants> (accessed on 1 June 2022).
24. Alm, E.; Broberg, E.K.; Connor, T.; Hodcroft, E.B.; Komissarov, A.B.; Maurer-Stroh, S.; Melidou, A.; Neher, R.A.; O’Toole, Á.; Pereyaslov, D.; et al. Geographical and Temporal Distribution of SARS-CoV-2 Clades in the WHO European Region, January to June 2020. *Euro Surveill.* **2020**, *25*, 2001410. [[CrossRef](#)] [[PubMed](#)]
25. Jung, C.; Kmiec, D.; Koepke, L.; Zech, F.; Jacob, T.; Sparrer, K.M.J.; Kirchhoff, F. Omicron: What Makes the Latest SARS-CoV-2 Variant of Concern So Concerning? *J. Virol.* **2022**, *96*, e0207721. [[CrossRef](#)] [[PubMed](#)]
26. Crits-Christoph, A.; Kantor, R.S.; Olm, M.R.; Whitney, O.N.; Al-Shayeb, B.; Lou, Y.C.; Flamholz, A.; Kennedy, L.C.; Greenwald, H.; Hinkle, A.; et al. Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants. *mBio* **2021**, *12*, 1110–1128. [[CrossRef](#)] [[PubMed](#)]
27. Fontenele, R.S.; Kraberger, S.; Hadfield, J.; Driver, E.M.; Bowes, D.; Holland, L.A.; Faleye, T.O.C.; Adhikari, S.; Kumar, R.; Inchausti, R.; et al. High-Throughput Sequencing of SARS-CoV-2 in Wastewater Provides Insights into Circulating Variants. *Water Res.* **2021**, *205*, 117710. [[CrossRef](#)] [[PubMed](#)]
28. Wang, R.; Chen, J.; Gao, K.; Hozumi, Y.; Yin, C.; Wei, G.-W. Analysis of SARS-CoV-2 Mutations in the United States Suggests Presence of Four Substrains and Novel Variants. *Commun. Biol.* **2021**, *4*, 228.
29. Rouchka, E.C.; Chariker, J.H.; Chung, D. Variant Analysis of 1040 SARS-CoV-2 Genomes. *PLoS ONE* **2020**, *15*, e0241535. [[CrossRef](#)] [[PubMed](#)]
30. Karthikeyan, S.; Levy, J.I.; De Hoff, P.; Humphrey, G.; Birmingham, A.; Jepsen, K.; Farmer, S.; Tubb, H.M.; Valles, T.; Tribelhorn, C.E.; et al. Wastewater Sequencing Reveals Early Cryptic SARS-CoV-2 Variant Transmission. *Nature* **2022**, *609*, 101–108. [[CrossRef](#)]
31. Smyth, D.S.; Trujillo, M.; Gregory, D.A.; Cheung, K.; Gao, A.; Graham, M.; Guan, Y.; Guldenpfennig, C.; Hoxie, I.; Kannoly, S.; et al. Tracking Cryptic SARS-CoV-2 Lineages Detected in NYC Wastewater. *Nat. Commun.* **2022**, *13*, 635. [[CrossRef](#)]
32. Vo, V.; Tillett, R.L.; Papp, K.; Shen, S.; Gu, R.; Gorzalski, A.; Siao, D.; Markland, R.; Chang, C.-L.; Baker, H.; et al. Use of Wastewater Surveillance for Early Detection of Alpha and Epsilon SARS-CoV-2 Variants of Concern and Estimation of Overall COVID-19 Infection Burden. *Sci. Total Environ.* **2022**, *835*, 155410. [[CrossRef](#)] [[PubMed](#)]
33. O’Reilly, K.M.; Allen, D.J.; Fine, P.; Asghar, H. The Challenges of Informative Wastewater Sampling for SARS-CoV-2 Must Be Met: Lessons from Polio Eradication. *Lancet Microbe* **2020**, *1*, e189–e190. [[CrossRef](#)] [[PubMed](#)]
34. Wilder, M.L.; Middleton, F.; Larsen, D.A.; Du, Q.; Fenty, A.; Zeng, T.; Insaf, T.; Kilaru, P.; Collins, M.; Kmush, B.; et al. Co-Quantification of CrAssphage Increases Confidence in Wastewater-Based Epidemiology for SARS-CoV-2 in Low Prevalence Areas. *Water Res. X* **2021**, *11*, 100100. [[CrossRef](#)] [[PubMed](#)]
35. Steele, J.A.; Zimmer-Faust, A.G.; Griffith, J.F.; Weisberg, S.B. Sources of Variability in Methods for Processing, Storing, and Concentrating SARS-CoV-2 in Influent from Urban Wastewater Treatment Plants. *bioRxiv* **2021**. [[CrossRef](#)]
36. Newton, R.J.; McClary, J.S. The Flux and Impact of Wastewater Infrastructure Microorganisms on Human and Ecosystem Health. *Curr. Opin. Biotechnol.* **2019**, *57*, 145–150. [[CrossRef](#)] [[PubMed](#)]
37. Achak, M.; Alaoui Bakri, S.; Chhiti, Y.; M’hamdi Alaoui, F.E.; Barka, N.; Boumya, W. SARS-CoV-2 in Hospital Wastewater during Outbreak of COVID-19: A Review on Detection, Survival and Disinfection Technologies. *Sci. Total Environ.* **2021**, *761*, 143192. [[CrossRef](#)]

38. Baaijens, J.A.; Zulli, A.; Ott, I.M.; Petrone, M.E.; Alpert, T.; Fauver, J.R.; Kalinich, C.C.; Vogels, C.B.F.; Breban, M.I.; Duvallet, C.; et al. Variant Abundance Estimation for SARS-CoV-2 in Wastewater Using RNA-Seq Quantification. *medRxiv* **2021**. [CrossRef]
39. Amman, F.; Markt, R.; Endler, L.; Hupfaut, S.; Agerer, B.; Schedl, A.; Richter, L.; Zechmeister, M.; Bicher, M.; Heiler, G.; et al. National-Scale Surveillance of Emerging SARS-CoV-2 Variants in Wastewater. *medRxiv* **2022**. [CrossRef]
40. Bivins, A.; Greaves, J.; Fischer, R.; Yinda, K.C.; Ahmed, W.; Kitajima, M.; Munster, V.J.; Bibby, K. Persistence of SARS-CoV-2 in Water and Wastewater. *Environ. Sci. Technol. Lett.* **2020**, *7*, 937–942. [CrossRef]
41. Ahmed, W.; Simpson, S.L.; Bertsch, P.M.; Bibby, K.; Bivins, A.; Blackall, L.L.; Bofill-Mas, S.; Bosch, A.; Brandão, J.; Choi, P.M.; et al. Minimizing Errors in RT-PCR Detection and Quantification of SARS-CoV-2 RNA for Wastewater Surveillance. *Sci. Total Environ.* **2022**, *805*, 149877. [CrossRef]
42. Philo, S.E.; Keim, E.K.; Swanstrom, R.; Ong, A.Q.W.; Burnor, E.A.; Kossik, A.L.; Harrison, J.C.; Demeke, B.A.; Zhou, N.A.; Beck, N.K.; et al. A Comparison of SARS-CoV-2 Wastewater Concentration Methods for Environmental Surveillance. *Sci. Total Environ.* **2021**, *760*, 144215. [CrossRef] [PubMed]
43. Rothman, J.A.; Saghir, A.; Chung, S.-A.; Boyajian, N.; Dinh, T.; Kim, J.; Oval, J.; Sharavanan, V.; York, C.; Zimmer-Faust, A.G.; et al. Longitudinal Metatranscriptomic Sequencing of Southern California Wastewater Representing 16 Million People from August 2020–21 Reveals Widespread Transcription of Antibiotic Resistance Genes. *Water Res.* **2022**, *229*, 119421. [CrossRef] [PubMed]
44. Bushnell, B. *BBTools Software Package*, 2014. Available online: <https://sourceforge.net/projects/bbmap/> (accessed on 25 March 2024).
45. *Picard Toolkit*; Broad Institute: Cambridge, MA, USA, 2019.
46. Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef] [PubMed]
47. Wood, D.E.; Lu, J.; Langmead, B. Improved Metagenomic Analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 257. [CrossRef] [PubMed]
48. Lu, J.; Breitwieser, F.P.; Thielen, P.; Salzberg, S.L. Bracken: Estimating Species Abundance in Metagenomics Data. *PeerJ Comput. Sci.* **2017**, *2017*, e104. [CrossRef]
49. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
50. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2009.
51. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; et al. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* **2020**, *579*, 265–269. [CrossRef] [PubMed]
52. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
53. Grubaugh, N.D.; Gangavarapu, K.; Quick, J.; Matteson, N.L.; De Jesus, J.G.; Main, B.J.; Tan, A.L.; Paul, L.M.; Brackney, D.E.; Grewal, S.; et al. An Amplicon-Based Sequencing Framework for Accurately Measuring Intra-host Virus Diversity Using PrimalSeq and iVar. *Genome Biol.* **2019**, *20*, 8. [CrossRef] [PubMed]
54. Turakhia, Y.; Thornlow, B.; Hinrichs, A.S.; De Maio, N.; Gozashti, L.; Lanfear, R.; Haussler, D.; Corbett-Detig, R. Ultrafast Sample Placement on Existing TRees (USHER) Enables Real-Time Phylogenetics for the SARS-CoV-2 Pandemic. *Nat. Genet.* **2021**, *53*, 809–816. [CrossRef]
55. Elbe, S.; Buckland-Merrett, G. Data, Disease and Diplomacy: GISAID’s Innovative Contribution to Global Health. *Glob. Chall.* **2017**, *1*, 33–46. [CrossRef]
56. California Health and Human Services Agency. *COVID-19 Variant Data—California Health and Human Services Open Data Portal*; California Health and Human Services Agency: Sacramento, CA, USA, 2023.
57. Mallick, H.; Rahnavard, A.; McIver, L.J.; Ma, S.; Zhang, Y.; Nguyen, L.H.; Tickle, T.L.; Weingart, G.; Ren, B.; Schwager, E.H.; et al. Multivariable Association Discovery in Population-Scale Meta-Omics Studies. *PLoS Comput. Biol.* **2021**, *17*, e1009442. [CrossRef] [PubMed]
58. Oksanen, J.; Blanchet, F.G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P.R.; O’Hara, R.B.; Simpson, G.L.; Solymos, P.; et al. *Vegan: Community Ecology Package*, 2017. Available online: <https://cran.r-project.org/web/packages/vegan/vegan.pdf> (accessed on 25 March 2024).
59. Kuznetsova, A.; Brockhoff, P.B.; Christensen, R.H.B. {lmerTest} Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw.* **2017**, *82*, 1–26. [CrossRef]
60. Slowikowski, K. *Ggrepel: Automatically Position Non-Overlapping Text Labels with “Ggplot2”*, R package version 0. 8. 0. 2018. Available online: <https://rdr.io/cran/ggrepel/> (accessed on 25 March 2024).
61. Nowosad, J. *Rcartocolor: ‘CARTOCOLOR’ Palettes*, 2018. Available online: <https://jakubnowosad.com/rcartocolor/> (accessed on 25 March 2024).
62. Pedersen, T.L. *Patchwork: The Composer of Plots*, R Package Version 1.2.0.9000. 2020. Available online: <https://rdr.io/cran/ggrepel/> (accessed on 25 March 2024).
63. Schumann, V.-F.; de Castro Cuadrat, R.R.; Wyler, E.; Wurmus, R.; Deter, A.; Quedenau, C.; Dohmen, J.; Foxel, M.; Borodina, T.; Blume, A.; et al. SARS-CoV-2 Infection Dynamics Revealed by Wastewater Sequencing Analysis and Deconvolution. *Sci. Total Environ.* **2022**, *853*, 158931. [CrossRef] [PubMed]
64. Brunner, F.S.; Brown, M.R.; Bassano, I.; Denise, H.; Khalifa, M.S.; Wade, M.J.; van Aerle, R.; Kevill, J.L.; Jones, D.L.; Farkas, K.; et al. City-Wide Wastewater Genomic Surveillance through the Successive Emergence of SARS-CoV-2 Alpha and Delta Variants. *Water Res.* **2022**, *226*, 119306. [CrossRef] [PubMed]

65. Fang, S.; Li, K.; Shen, J.; Liu, S.; Liu, J.; Yang, L.; Hu, C.-D.; Wan, J. GESS: A Database of Global Evaluation of SARS-CoV-2/HCoV-19 Sequences. *Nucleic Acids Res.* **2021**, *49*, D706–D714. [[CrossRef](#)] [[PubMed](#)]
66. Chen, A.T.; Altschuler, K.; Zhan, S.H.; Chan, Y.A.; Deverman, B.E. COVID-19 CG Enables SARS-CoV-2 Mutation and Lineage Tracking by Locations and Dates of Interest. *eLife* **2021**, *10*, e63409. [[CrossRef](#)] [[PubMed](#)]
67. Rios, G.; Lacoux, C.; Leclercq, V.; Diamant, A.; Lebrigand, K.; Lazuka, A.; Soyeux, E.; Lacroix, S.; Fassy, J.; Couesnon, A.; et al. Monitoring SARS-CoV-2 Variants Alterations in Nice Neighborhoods by Wastewater Nanopore Sequencing. *Lancet Reg. Health Eur.* **2021**, *10*, 100202. [[CrossRef](#)] [[PubMed](#)]
68. Nemudryi, A.; Nemudraia, A.; Wiegand, T.; Surya, K.; Buyukyoruk, M.; Cicha, C.; Vanderwood, K.K.; Wilkinson, R.; Wiedenheft, B. Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater. *Cell Rep. Med.* **2020**, *1*, 100098. [[CrossRef](#)] [[PubMed](#)]
69. Koskela von Sydow, A.; Lindqvist, C.M.; Asghar, N.; Johansson, M.; Sundqvist, M.; Mölling, P.; Stenmark, B. Comparison of SARS-CoV-2 Whole Genome Sequencing Using Tiled Amplicon Enrichment and Bait Hybridization. *Sci. Rep.* **2023**, *13*, 6461. [[CrossRef](#)]
70. Vanaerschot, M.; Mann, S.A.; Webber, J.T.; Kamm, J.; Bell, S.M.; Bell, J.; Hong, S.N.; Nguyen, M.P.; Chan, L.Y.; Bhatt, K.D.; et al. Identification of a Polymorphism in the N Gene of SARS-CoV-2 That Adversely Impacts Detection by Reverse Transcription-PCR. *J. Clin. Microbiol.* **2020**, *59*, e02369-20. [[CrossRef](#)]
71. Wu, F.; Zhang, J.; Xiao, A.; Gu, X.; Lee, W.L.; Armas, F.; Kauffman, K.; Hanage, W.; Matus, M.; Ghaeli, N.; et al. SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases. *mSystems* **2020**, *5*, e00614-20. [[CrossRef](#)] [[PubMed](#)]
72. Rothman, J.A.; Loveless, T.B.; Griffith, M.L.; Steele, J.A.; Griffith, J.F.; Whiteson, K.L. Metagenomics of Wastewater Influent from Southern California Wastewater Treatment Facilities in the Era of COVID-19. *Microbiol. Resour. Announc.* **2020**, *9*, 19–21. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.