Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Direct comparison of RT-ddPCR and targeted amplicon sequencing for SARS-CoV-2 mutation monitoring in wastewater

Esther G. Lou\textsuperscript{a}, Nicolae Sapoval\textsuperscript{b}, Camille McCall\textsuperscript{a}, Lauren Bauhs\textsuperscript{a}, Russell Carlson-Stadler\textsuperscript{a}, Prashant Kalvapalle\textsuperscript{c}, Yanlai Lai\textsuperscript{d}, Kyle Palmer\textsuperscript{a}, Ryker Penn\textsuperscript{d}, Whitney Rich\textsuperscript{a}, Madeline Wolken\textsuperscript{a}, Pamela Brown\textsuperscript{d}, Katherine B. Ensor\textsuperscript{e}, Loren Hopkins\textsuperscript{d,e}, Todd J. Treangen\textsuperscript{b}, Lauren B. Stadler\textsuperscript{a,⁎}

\textsuperscript{a} Department of Civil and Environmental Engineering, Rice University, 6100 Main Street, MS 519, Houston, TX 77005, United States of America
\textsuperscript{b} Department of Computer Science, Rice University, 6100 Main Street, Houston, TX 77005, United States of America
\textsuperscript{c} Systems, Synthetic, and Physical Biology Graduate Program, Rice University, 6100 Main Street, Houston, TX 77005, United States of America
\textsuperscript{d} Houston Health Department, 8000 N. Stadium Dr., Houston, TX 77054, United States of America
\textsuperscript{e} Department of Statistics, Rice University, 6100 Main Street MS 138, Houston, TX 77005, United States of America

HIGHLIGHTS

- 547 wastewater samples were analyzed using RT-ddPCR and ARTIC-based targeted amplicon sequencing in parallel
- RT-ddPCR is more sensitive than ARTIC-based amplicon sequencing for detection and quantification on five mutations
- Neither AF nor reads containing mutations via sequencing can be directly used to estimate the mutation concentration
- Sequencing detections were positively influenced by sequencing depth, SARS-CoV-2 concentration, and mutation concentration
- RT-ddPCR should be applied for quantitative analysis and is preferable for early variant detection if targets are known

GRAPHICAL ABSTRACT

Abstract

Over the course of the COVID-19 pandemic, variants of SARS-CoV-2 have emerged that are more contagious and more likely to cause breakthrough infections. Targeted amplicon sequencing approach is a gold standard for identification and analysis of variants. However, when applied to environmental samples such as wastewater, it remains unclear how sensitive this method is for detecting variant-associated mutations in environmental samples. Here we directly compare a targeted amplicon sequencing approach (using ARTIC v3; hereafter referred to as sequencing) with RT-ddPCR quantification for the detection of five mutations that are characteristic of variants of concern (VoCs) in wastewater samples. In total, 547 wastewater samples were analyzed using both methods in parallel. When we observed positive mutation detections by RT-ddPCR, 42.6% of the detection events were missed by sequencing, due to negative detection or the limited read coverage at the mutation position. Further, when sequencing reported negative or depth-limited mutation detections, 26.7% of those events were instead positive detections by RT-ddPCR, highlighting the relatively poor sensitivity of
1. Introduction

Over the course of the COVID-19 pandemic, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has evolved and numerous lineages have emerged that are more transmissible, cause more severe disease, and/or are better at escaping the immune response system (Garcia-Beltran et al., 2021; Harvey et al., 2021; Li et al., 2020b). Tracking the emergence and spread of these variants of concern (VoCs) and variants of interest (Vols) has become critical to public health response and mitigation strategies for stopping the spread of SARS-CoV-2. Wastewater-based epidemiology (WBE) is one prominent approach that has been adopted by public health departments and water utilities to track infection dynamics in communities by quantifying the amount of SARS-CoV-2 RNA in wastewater samples (Ahmed et al., 2020; Arora et al., 2020). WBE can also be used for monitoring VoCs and Vols in communities (Bar-Or et al., 2021; Fontenele et al., 2021; Heijnen et al., 2021).

Variant identification in wastewater samples is challenging because the viral genomes are highly fragmented, dilute, and comprised of mixtures of circulating variants. The most common methods used for wastewater variant screening are: (1) specifying specific characteristic mutations via RT-qPCR or RT-ddPCR (Ciesielski et al., 2021; Heijnen et al., 2021); and (2) enriching and sequencing SARS-CoV-2 genomes in wastewater (Swift et al., 2021; Tyson et al., 2020). Underlying both approaches is the ability to identify and quantify characteristic mutations that define variants. RT-qPCR is regarded as a gold standard method for routine wastewater surveillance (Alygizakis et al., 2021; Rahman et al., 2021; Van Poelvoorde et al., 2021). Compared to RT-qPCR, RT-ddPCR has a superior detection sensitivity (Ahmed et al., 2022; Ciesielski et al., 2021; Flood et al., 2021) and is less sensitive to inhibitors present in wastewater (Cao et al., 2015; Ciesielski et al., 2021). However, PCR based methods are limited in that the mutations must be known ahead of time for primer and probe design. In addition, in practice, they are limited by the number of targets that can be multiplexed per reaction, and it is difficult to delineate all variants present in the sample by only targeting a few mutations. Next generation sequencing (NGS) enables comprehensive screening of all potential mutations without any prior knowledge, and thus has been frequently applied for characterizing pathogens and viruses (Greninger et al., 2015; Yang et al., 2011). The unbiased, non-targeted metagenomics sequencing approaches often require high- or ultra-high coverage in order to obtain enough target sequences (Chiara et al., 2021). On the other hand, targeted sequencing approaches using an enrichment step during library preparation maximize the detection of viruses effectively (Deng et al., 2020). For SARS-CoV-2 genome enrichment, multiplex tiling PCR and oligonucleotide capture are the most frequently implemented methods, both demonstrating great performance in terms of genome coverage (Doddapaneni et al., 2021; Tyson et al., 2020) and mutation detection (Bar-Or et al., 2021; Crits-Christoph et al., 2021). Targeted amplicon sequencing (i.e., multiplex tiling PCR coupled with amplicon sequencing) is considered the lower-cost and faster approach (Chiara et al., 2021; Lin et al., 2021). For example, ARTIC Network panels (https://artic.network) are commonly used by laboratories globally to characterize SARS-CoV-2 present in clinical samples (Charre et al., 2020; Li et al., 2020a; Mboowa et al., 2021). However, targeted amplicon sequencing is frequently limited by coverage and/or quality dropout due to amplification bias and primer knockout by mutations that happen to occur at the priming regions (Davis et al., 2021).

Several studies used amplicon-based sequencing of SARS-CoV-2 from wastewater samples to estimate the prevalence of SARS-CoV-2 variants in the communities (Bar-Or et al., 2021; Layton et al., 2021; Otero et al., 2021). One study used AF (allele frequency) of VoC-associated mutations detected in wastewater sample to estimate the relative abundances of different lineages circulating in a community (Ellmen et al., 2021). However, it is unclear how quantitative measurement AF are that are generated from wastewater genomes via this approach. Thus, a direct comparison between targeted amplicon sequencing and RT-ddPCR (or RT-qPCR) for mutation detection and quantification using wastewater samples is needed. In this study, we quantified five unique mutations using RT-ddPCR and performed targeted amplicon sequencing (ARTIC v3 based) of SARS-CoV-2 in parallel on 547 wastewater samples. We compare the consistency in the approaches in terms of (1) detection vs. no detection; and (2) quantitative information generated by each method. In addition, we evaluated the impact of mutation concentration, single base coverage at the mutation position, the overall SARS-CoV-2 concentrations, and SARS-CoV-2 genome coverage on mutation detection via targeted amplicon sequencing.

2. Materials and methods

2.1. Wastewater sample collection, concentration, and RNA extraction

We collected weekly wastewater samples from 39 wastewater treatment plants (WWTPs) in Houston covering a service area of approximately 580 mile² and serving over 2.3 million people. Time-weighted composite samples of raw wastewater were collected every 1 h for 24 h from the influent of the WWTPs. The sampling campaign was conducted during two separate periods of time (Phase I and Phase II). Phase I spanned from February 23, 2021 to April 12, 2021 when Alpha (specifically, B.1.1.7) was the dominant SARS-CoV-2 VoC circulating in Texas (GISAID, https://www.gisaid.org/). Phase II covered from May 24, 2021 to July 12, 2021 when the Delta variant became prevalent, displacing Alpha as reflected by variant confirmed cases in Texas (GISAID, https://www.gisaid.org/). In total, 249 and 298 samples were analyzed during Phase I and Phase II, respectively. SARS-CoV-2 was concentrated in wastewater samples using an electron-activated filtration method as previously described (LaTurner et al., 2021). RNA extraction was performed using a Chemagic™ Viral DNA/RNA 300 Kit H96 (Chemagic, CMG-1433, PerkinElmer) with the PerkinElmer viral RNA/DNA purification protocol and reagents. Finally, 10 μL of sample extract was used for each RT-ddPCR reaction, and 11 μL of sample extract was used for sequencing library preparation. Detailed concentration procedures, RNA extraction procedures, concentration factors (Table S1), and associated quality control measurements are provided in the Supplementary materials following the Environmental Microbiology Minimum Information (EMMI) Guidelines (Borchardt et al., 2021).

2.2. RT-ddPCR quantification of SARS-CoV-2 N1, N2 genes, and five characteristic mutations

RT-ddPCR was performed on a QX200 AutoDG Droplet Digital PCR System (Bio-Rad) and a C1000 Thermal Cycler (Bio-Rad) in 96-well optical plates. SARS-CoV-2 N1 and N2 gene targets were quantified in wastewater samples as previously described (LaTurner et al., 2021). Five mutations, namely S:DEL69/70, S:N501Y, S:E484K, S:K417T, S:L452R, were quantified via RT-ddPCR. GT Molecular kits were used for RT-ddPCR quantification of mutations (kit information provided in Table S2). S:DEL69/70 and S:N501Y, two characteristic mutations associated with the Alpha lineage B.1.1.7, were quantified during Phase I, and S:L452R, S:K417T and S:E484K were quantified during Phase II. The latter three mutations were selected due to their reducing SARS-CoV-2 susceptibility to convalescent and vaccine-elicited sera and mAbs, and their emergence in newly evolved SARS-CoV-2 strains (Jangra et al., 2021; Wilhelm et al., 2021). These SARS-CoV-2 mutations were quantified using one-step RT-ddPCR assays according to the manufacturer’s protocol (GT Molecular). A detailed
description of the methods, including droplet thresholding and limit of detection (LOD) are described in the Supplementary materials (Section 1.4, Table S3 – S5). For all targets (N1, N2, and the five mutations), positive detection (+) was defined as above the LOD, and a negative detection (−) was defined as below the LOD. RT-ddPCR analysis was used to generate: (1) the concentration of the mutation in copies/L-wastewater, and (2) the fraction of SARS-CoV-2 containing the mutation, which was calculated by normalizing the copies of the mutation by the sum of the copies of the mutation and the wild-type.

2.3. Amplicon-based sequencing using ARTIC v3 and data analysis

cDNA was generated using 11 μL RNA extract via reverse transcription using the Superscript IV first-strand synthesis system (ThermoFisher Scientific, 18,091,050) following the manufacturer’s protocol. SARS-CoV-2 genome enrichment via multiplexing PCR was conducted using ARTIC v3 protocol (Tyson et al., 2020). Illumina DNA Prep kit with the manufacturer’s manual (DNA Flex) were applied for amplicon tagmentation and flex amplification, followed by library clean-up. Each sample library was then quantitated, normalized, pooled, and diluted to 6 pM. Finally, sequencing was performed on an Illumina MiSeq instrument using MiSeq Reagent Kit v2 (300-cycles, MS-103-2002) following a 151 + 10 + 10 + 151 cycling recipe. Sequencing read trimming with BBduk (Bushnell, 2014) was conducted, followed by read mapping with BWA-MEM (H. Li, 2013). Read mapping result was sorted by samtools (H. Li et al., 2009), and primer locations were soft-clipped using iVar (Grubaugh et al., 2019). Finally, variant calls were performed with respect to the Wuhan reference sequence (NC_045512.2) using LoFreq (Wilm et al., 2012).

To compare targeted amplicon sequencing (hereafter referred to as sequencing) with RT-ddPCR, we first calculated the sequencing read coverage for each target mutation in each sample by averaging the single base coverage across the target region used for quantification in RT-ddPCR. For example, for the N1 and N2 RT-ddPCR assays, the CDC 2019-nCoV_N1 probe and 2019-nCoV_N2 probe were applied to RT-ddPCR assays in this study. These probes align to nt 28318 – 28332 and nt 29188 – 29210, respectively (the nt coordinates correspond to the Wuhan reference sequence NC_045512.2). Accordingly, sequencing mapped reads for N1 and N2 were checked at each single base position from nt 28318 to 28322 (N1 region, containing 15 positions) and from nt 29188 to 29210 (N2 region, containing 23 positions), respectively. A positive detection (+) was called for N1 and N2 if (1) the single base coverage at each nt position across the target region (nt 28318 – 28332 for N1 and nt 29188 – 29210 for N2) was at least 1×, and (2) the average single base coverage across the target region was at least 20×; otherwise, a non-detect (ND) was called. For the five mutations, if there was at least 20× read depth at the position that corresponded to the target mutation, a positive detection (+) was called if any reads containing that mutation were observed. A negative detection for sequencing (−) was defined as no reads containing the target mutation were observed, and there was at least 20× read depth at the mutation location. Finally, for sequencing if less than 20 reads mapped to the mutation position, we defined this as “depth limited (DL)”. We used a 20× coverage threshold for sequencing analysis based on previous studies that applied tiled PCR and short-read sequencing (Illumina) for SARS-CoV-2 wastewater analysis (Baaijens et al., 2021; Fontenele et al., 2021).

2.4. Statistical analysis

Welch two sample t-test was applied to compare datasets. Spearman rank correlation analysis was used to study the associations between quantitative results generated by RT-ddPCR and sequencing. For each target mutation, we used Spearman rank correlation to assess the correlations between (1) the mutation concentration as determined by RT-ddPCR (copies/L-wastewater) and the number of reads containing the mutation as determined by sequencing; and (2) the fraction of SARS-CoV-2 containing the target mutation (target mutation concentration/total SARS-CoV-2 concentration as determined by RT-ddPCR) and the AF of the mutation as determined by sequencing. Strength of correlations were identified based on Spearman’s correlation coefficient Rho (R).

3. Results and discussion

3.1. RT-ddPCR was more sensitive than sequencing for mutation detection

547 wastewater samples (249 samples for Phase I, 298 samples for Phase II) were analyzed using both RT-ddPCR (targeting N1, N2, and five mutations) and sequencing. The wastewater concentrations of SARS-CoV-2 (determined by the average of N1 and N2 concentrations) were significantly higher during Phase I than during Phase II (Fig. S1, p < 0.001). For sequencing, an average fraction of 0.617 (std: 0.222) of the reads in Phase I, and an average fraction of 0.298 (std: 0.227) of the reads in Phase II mapped to the SARS-CoV-2 Wuhan reference. For the 547 wastewater samples, sequencing generated an average of 155,583 ± 249,909 reads (Phase I: 295,938 ± 325,644; Phase II: 46,784 ± 41,349) that mapped to the reference genome. The average single base coverage across the entire SARS-CoV-2 genome was 973 reads per base (range: 0.03 to 4444). The average breadth of coverage, defined as the percentage of genome bases sequenced per sample, was 66.7% (range: 1.4% to 99.9%; sequencing read statistics detailed in Table S6). Additional information on all samples analyzed, including their detections by RT-ddPCR and sequencing for each target are detailed in Fig. S2.

To compare RT-ddPCR and sequencing, we categorized detection events into different scenarios. There were four possible scenarios for N1 and N2 detections: positive detections by both RT-ddPCR and sequencing (+/ +), positive detection by RT-ddPCR and not detected by sequencing (+/ −), negative detection by RT-ddPCR and positive detection by sequencing (−/ +), and negative detection by RT-ddPCR and not detected by sequencing (−/ −). For the five target mutations, there were six possible scenarios: positive detections by both RT-ddPCR and sequencing (+/ +), positive detection by RT-ddPCR and negative detection by sequencing (+/ −), positive detection by RT-ddPCR and depth limited (single base coverage < 20× at the mutation position) for sequencing (+/ DL), negative detections by both methods (−/ −), negative detection (−/ −), positive detection by RT-ddPCR and depth limited (single base coverage < 20× at the mutation position) for sequencing (+/ DL), negative detection by both methods (−/ −), and negative detection (−/ −). The bars in the middle group are based on mutation detection events by RT-ddPCR and sequencing, the bars on the left and right group are based on the horizontal axis representing the total number of mutations detected by RT-ddPCR and sequencing. The bars on the left and right group are based on the horizontal axis representing the total number of mutations detected by RT-ddPCR and sequencing.
by RT-ddPCR and positive detection by sequencing (−/+), and negative detection by RT-ddPCR and depth limited by sequencing (−/DL). We first assessed the relationship between N1 and N2 detections and mutation detections. Fig. 1 shows each detection event (center bars) and their corresponding N1 and N2 detections. As expected, almost all samples with RT-ddPCR and/or sequencing positive detections for mutations ([+/+], (+/−), (+/DL); (−/+)) were also positive for N1 and N2 (mainly (+/+) and (+/ND)).

Among all 1094 N1 and N2 detection events in 547 samples, 67.6% had consistent detections for RT-ddPCR and sequencing ([(+/+), (+/−), (+/DL); (−/+)] and (−/−)). For RT-ddPCR, we observed very consistent N1 and N2 detections for the vast majority (97.1%) of samples, with N1 and N2 double positive and N1 and N2 double ND events for 531 of the 547 samples. In contrast, N1 and N2 detections via sequencing reported consistent detections for only 403 samples (73.7% of total samples). The inconsistency of sequencing in N1 and N2 detections was likely due to the significantly lower read depth at the N1 region than at the N2 region (p < 0.0001). Overall, RT-ddPCR detected 499 N1 and N2 double positive events whereas sequencing only detected 265. These results indicate that sequencing detection is less sensitive than RT-ddPCR detection when focusing on commonly targeted N-gene regions of the SARS-CoV-2 genome.

Unsurprisingly, we also found that sequencing was less sensitive than RT-ddPCR for detecting target mutations. We compared 1354 possible mutation detections using both RT-ddPCR and sequencing (Table S2, Fig. S2), in terms of their consistency in calling the presence or absence of mutations (Fig. 2). Among all 1354 detection events, only 39.6% represented consistent detections for both methods ([(+/+), (−/−)] (Fig. 2). The inconsistency was mainly attributed to scenario (−/DL) where mutation detections were confirmed as negative by RT-ddPCR and were limited by the depth at the mutation positions for sequencing. The scenario (−/DL) alone accounted for 39.0% of all detection events, suggesting sequencing is very likely to miss a negative detection for a mutation. The large number of DL events by sequencing largely occurred during Phase II for the detection of mutations S:E484K, S:K417T and S:L452R (Fig. 2). During Phase II, the concentrations of SARS-CoV-2 were significantly lower than SARS-CoV-2 concentrations during Phase I when the samples were assayed for S:DEL69/70 and S:N501Y (Fig. S1, p < 0.0001). Phase II samples also had lower average single base coverage across the genome by sequencing as compared to Phase I samples (Table S6, p < 0.0001).

When we observed positive detections by RT-ddPCR ([+/+), (+−), and (+/DL)], 42.6% of the detection events ([(+/+), (+/−), and (+/DL)] were missed by sequencing. When sequencing reported negative or depth limited detections ([+/−],[+/DL], (−/−), and (−/DL)], 26.7% of those events were detected as positive by RT-ddPCR ([+/+), (+/−), and (+/+)] (Fig. 3). Two findings indicate that sequencing was less sensitive than RT-ddPCR for mutation detection. The majority (96.1%) of RT-ddPCR negative detections ([+/−], (−/−), and (−/DL)] were also consistently called as negative or depth-limited by sequencing (−/−), (−/DL). The concentration of the target mutation in a sample is, in part, a function of the number of individuals infected with a SARS-CoV-2 variant containing that mutation. Thus, it would be low in the early stages of a variant outbreak, in which case sequencing may not be as sensitive as RT-ddPCR for early detection.

3.2. Impact of mutation concentration, single base coverage at the mutation position, SARS-CoV-2 concentration, and the average single base coverage across the entire genome on mutation detection

To understand when the inconsistencies between sequencing and RT-ddPCR were more likely to occur, we evaluated the impact of the mutation concentration (as determined by RT-ddPCR) and the average single base coverage at the mutation position (as determined by sequencing) on mutation detection events. First, we compared mutation concentrations in the sample as determined by RT-ddPCR for events where both methods reported positive detections (+/+) to events where RT-ddPCR reported positive and sequencing reported negative or depth limited detections ([+/−], (+/DL)] (Fig. 3a). Then, we compared the single base coverage at the mutation position for these scenarios as determined by sequencing (Fig. 3b). Results revealed samples with (+/−) and (+/DL) events had significantly lower mutation concentrations and single base coverage than those with (+/+) events (t-test, p < 0.001; Fig. 3).

In addition, we evaluated the impact of the sample SARS-CoV-2 concentration (average of N1 and N2 concentrations as determined by RT-ddPCR) and the average single base coverage (read depth) across the entire SARS-CoV-2 genome (as determined by sequencing) on mutation detections (Fig. 4). We found that the SARS-CoV-2 concentration and the average read depth across the entire genome were significantly higher for RT-ddPCR positive detections ([(+/+), (+/−), and (+/DL)], n = 612) than for samples with RT-ddPCR negative detections ([+/−], (−/−), and (−/DL)], n = 742; Fig. 4). Further, we observed significantly higher SARS-CoV-2 concentrations and the average read depth across the entire genome in samples with sequencing positive detections ([(+/+), (+/−), n = 342]) than samples with sequencing negative detections ([(+/−), (−/−), n = 296] or DL [(+/DL), (−/DL), n = 716]; Fig. 4).

In addition to SARS-CoV-2 concentration and read depth impacting sequencing detection, multiplexing PCR may also introduce amplification bias due to dimer formation and low Ta dropping-out (Coil et al., 2021; Tyson et al., 2020). This bias can result in an uneven coverage across the genome and impact the number of reads generated at different mutation positions, reducing the reliability of mutation calling. In this study, the two target B.1.1.7 characteristic mutations (S:N501Y and S:DEL69/70) are located in different regions of the SARS-CoV-2 genome, and they were amplified by different ARTIC v3 primers. The substitution S:N501Y (A23063T) was amplified by the primer pair ARTIC v3 F/R_76, while the deletion S:DEL69/70 (21764ATACATG) was amplified by primers ARTIC v3 F/R_72. We found that the average coverage of these two amplified regions were highly variable across samples (Fig. S3); and the average coverage across the region amplified by ARTIC v3 F/R_76 was significantly lower than the region amplified by ARTIC v3 F/R_72 (p < 0.05, n = 249). This amplification bias helps explain the significantly higher average single base coverage at nt position 21,764 (corresponding to S:DEL69/70) than at nt position 23,063 (corresponding to S:N501Y) (p < 0.05, n = 249). Correspondingly, sequencing exhibited more sensitive detection for S:DEL69/70 than S:N501Y, detecting 71.2% of S:DEL69/70 RT-ddPCR positive events and 60.2% of S:N501Y RT-ddPCR positive events, respectively (Fig. 2a).

With the aim of guiding wastewater-based SARS-CoV-2 monitoring in practice, we also attempted to identify whether there was a threshold
level of sequencing depth, or the total reads mapped to SARS-CoV-2 reference genome, above which sequencing called positive detections concomitantly with RT-ddPCR (Fig. S4). Two groups of samples belonging to scenarios (+/+) and (+/−) were used for this analysis, because there was no significant difference in SARS-CoV-2 concentration or in mutation concentration between them, ensuring that the sample sequencing depth is the only variable that determined mutation detection by sequencing. We were not able to identify a threshold level of reads that could differentiate the sequencing positive and negative detection events. In other words, there was no clear pattern in the number of reads for samples with sequencing positive versus negative detections (Fig. S4).

3.3. Allele frequency (AF) and the number of reads supporting the mutation from sequencing were not quantitative representations of the mutation concentration as determined by RT-ddPCR

We next asked if data generated from sequencing could be used to quantitatively estimate the proportion of SARS-CoV-2 that contained a target mutation. For each target mutation, we compared (1) the mutation concentration as determined by RT-ddPCR (copies/L-wastewater); and (2) the fraction of SARS-CoV-2 containing the target mutation (target mutation concentration/total SARS-CoV-2 concentration as determined by RT-ddPCR) to (3) the number of reads containing the mutation as determined by sequencing for all sequencing positive and negative detections (i.e., detections with at least 20 × read depth at the mutation position); and (4) the AF of the mutation from sequencing for all sequencing positive and negative detections. We observed weak positive correlations between mutation concentration and raw read counts with the mutation [(1) vs (3); \( R = 0.30, p < 0.0001 \)], and between fraction of SARS-CoV-2 with the mutation and AF [(2) vs (4); \( R = 0.32, p < 0.0001 \)]. No significant correlation was found between mutation concentration and AF [(1) vs (4); \( p = 0.12 \)], or between fraction of SARS-CoV-2 with the mutation and raw read counts with the mutation [(2) vs (3); \( p = 0.17 \)].

We further investigated the quantitative relationship (or lack thereof) between mutation levels of the two B.1.1.7 characteristic mutations, S:N501Y and S:DEL69/70, as determined by RT-ddPCR and sequencing.
during Phase I when B.1.1.7 was the dominate circulating variant in Houston. During this period, the RT-ddPCR-determined concentrations of S:N501Y and S:DEL69/70, as well as the fractions of SARS-CoV-2 containing S:N501Y and S:DEL69/70, were strongly positively correlated ($R = 0.95$, $p < 0.0001$ for S:N501Y and S:DEL69/70 concentrations; $R = 0.90$, $p < 0.0001$ for fractions of SARS-CoV-2 containing S:N501Y and S:DEL69/70; $n = 228$). These results indicate RT-ddPCR measurements of mutations present in a single, dominant VoC vary consistently with one another. However, when we look at the sequencing data, the correlation between the number of reads containing S:N501Y and S:DEL69/70, and the correlation between the AF values of S:N501Y and S:DEL69/70 were both weaker ($R = 0.21$, $p < 0.05$ for the number of reads containing S:N501Y and S:DEL69/70, $n = 121$; $R = 0.33$, $p < 0.001$ for AF values of S:N501Y and S:DEL69/70, $n = 121$). We also assessed associations between RT-ddPCR and sequencing measurements of N1 and N2 genes. Moderate positive correlations were observed between N1 concentration (via RT-ddPCR) and average single base coverage at N1 region (via sequencing, $n = 356$, $R = 0.46$, $p = 0.00011$), and between N2 concentration and average single base coverage at N2 region ($n = 311$, $R = 0.46$, $p = 0.00066$), respectively. Reasonably, N1 and N2 concentrations determined by RT-ddPCR exhibited a near perfect positive correlation ($R = 0.98$, $p < 0.0001$), reflecting the robustness of the CDC primers and probes for measuring SARS-CoV-2 levels in wastewater samples. On the other hand, the sequencing average coverage at N1 and N2 regions only displayed a moderate positive correlation with each other ($R = 0.46$, $p < 0.0001$). Previous studies that used RT-qPCR for N1 and N2 quantifications in wastewater samples found strong correlations between N1 and N2 signals, reporting correlation coefficients such as 0.952 (Sanjuán and Domingo-Calap, 2021) and above 0.99 (Ahmed et al., 2022). Studies that applied RT-ddPCR for N1 and N2 quantifications in wastewater samples also reported strong correlation coefficients above 0.85 (Feng et al., 2021) and above 0.90 (D’Aoust et al., 2021). These results highlight that AF values and the read counts as determined by sequencing for mutations may not vary consistently with one another, and thus are not appropriate for inferring VoC concentrations or relative abundances in wastewater samples.

3.4. Implications for WBE on SARS-CoV-2

RT-ddPCR or RT-qPCR should be applied for quantitative analyses due to the great sensitivity and consistency of detection. In addition, RT-ddPCR and RT-qPCR generally have much shorter result turnaround time compared to sequencing (Bloom et al., 2021), which is critical for real-time public health response. With knowledge of unique mutations associated with each VoC, it is possible to detect signatures of low levels of VOCs in wastewater samples that may contain a mixture of variants. Numerous studies have emerged that have successfully developed, validated, and applied RT-qPCR or RT-ddPCR assays for the detection of specific VOCs by targeting characteristic mutations. For instance, RT-qPCR assays have been developed to co-monitor B.1.1.7 and B.1.351 by tracking the trend of a B.1.1.7-specific mutation, D3, and a B.1.351-specific mutation, the deletion 242–244 (Erster et al., 2021; Yaniv et al., 2021). Recently, allele-specific and multiplex-compatible RT-qPCR assays targeting mutations T19R, D80A, K417N, T478K and E484Q for quantitative detection and discrimination of the Delta, Beta, Delta plus, Kappa and Beta variants in wastewater were developed and validated (Lee et al., 2021).

The lower detection sensitivity of sequencing (Fig. 2) can be attributed to 1) the low concentration of target mutations in the wastewater sample (Fig. 3a), 2) the lack of sufficient read depth at the mutation position (Fig. 3b), 3) low SARS-CoV-2 concentrations (Fig. 4b), and/or 4) inconsistent single base coverage across the SARS-CoV-2 genome (Fig. 4a, Fig. 5). The sensitivity of targeted amplicon sequencing can also be impacted by sample processing and primer choices for genome amplification. The form of SARS-CoV-2-RNA in wastewater has only been characterized in a limited number of studies, and likely exists in both intact and degraded forms (Canh et al., 2021; Wurtzer et al., 2021). The degraded form presents a challenge for amplification. For example, the ARTIC v3 primer scheme used in this study amplifies 400 bp regions of the genome, and thus may fail to amplify short, degraded RNA fragments. Further research is needed to optimize workflows including sample processing and tiled primer design for downstream sequencing and analysis using wastewater samples. For example, improvements in virus recovery and yields during wastewater sample concentration and viral RNA extraction could enhance sequencing sensitivity. In addition, little is known about the impact of concentration method on mutation detection via sequencing. Sensitivity could also be improved by developing multiplexing PCR schemes of higher amplification uniformity and efficiency (Itokawa et al., 2020), or optimizing library preparation protocols (Coil et al., 2021). PCR inhibition due to other constituents in wastewater is another factor that may impact the sensitivity of sequencing more than RT-ddPCR, as digital PCR is relatively resilient to PCR inhibition (Ahmed et al., 2022; Ciesielski et al., 2021). Another approach for increasing sequencing coverage at specific sites is to target a smaller region of the genome for amplification and sequencing, such as by sequencing only the spike protein region of SARS-CoV-2 instead of the whole genome. For example, the receptor binding domain (RBD) on the spike region of SARS-CoV-2, which is involved in the interactions with human angiotensin-converting enzyme-2 (ACE-2) receptor, can be sequenced instead of the entire genome for mutation or variant analysis (Gregory et al., 2021). The mutations in the RBD are associated with the severity of infection (i.e., ACE-2 binding affinity and virus entry to the host cells) (Andersen et al., 2020; Heald-Sargent and Gallagher, 2012) and potential antibody-escape affecting antigenicity (Harvey et al., 2021). In addition, many of the VoCs are defined by mutations to in the Spike region (Baijens et al., 2021). Furthermore, the spike region only accounts for approximately 12.8% of the total genome, therefore, may be a more efficient use of sequencing for mutation detection. However, this approach also suffers from amplification and sequencing challenges due to degraded RNA inherent to wastewater samples.

Despite its lower sensitivity and qualitative nature, sequencing still has a clear advantage of being more comprehensive, not limited by a priori knowledge of the target mutations, and enables the discovery of cryptic lineages (Smyth et al., 2022) and emerging lineages of concern (Sapoval et al., 2021). This can be critical for early detection of variants when the availability of primers and probes is limited or delayed due to supply chain challenges. In addition, sequencing data facilitates retrospective analyses, such as searching for specific mutations or collections of mutations present in samples collected prior to knowledge of the variants in communities (Johnson et al., 2022; La Rosa et al., 2021; Wilton et al., 2021). In practice, WBE systems can benefit from coupling sequencing with quantitative analyses such as RT-ddPCR or RT-qPCR to achieve a comprehensive picture of circulating mutations (using sequencing), and sensitive, quantitative information on variant-associated mutations (using RT-qPCR/RT-ddPCR).

4. Conclusion

For WBE work on SARS-CoV-2, sequencing technology has demonstrated irreplaceable advantages in efficient screening and the potentials to detect emerging or cryptic lineages. We performed RT-ddPCR and sequencing analyses in parallel on hundreds of wastewater samples for SARS-CoV-2 monitoring, with a specific focus on mutations associated with VOcs. This is the first study to directly compare mutation detection consistency between these two methods. Results first showed the significant agreement in detection sensitivity of RT-ddPCR in detecting five mutations as compared to amplicon-based sequencing. Secondly, quantitative results generated from sequencing, including allele frequency (AF) and single base coverage of specific mutations failed to reflect the concentrations of the corresponding mutations in wastewater, showing poor correlations with RT-ddPCR quantification results. Therefore, caution should be exercised in using sequencing for quantitative assessments of mutation abundance in wastewater samples. RT-ddPCR or RT-qPCR should be applied for quantitative analyses due to the great sensitivity and consistency of detection.
Appendix A. supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2022.155059.

References

Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O’Brien, J.W., Choi, P.M., Kitajima, M., Simpson, S.L., Li, J., Tischarke, B., Verhagen, R., Smith, W.J.M., Zaugg, J., McArthur, C., Chiu, C.Y., 2020. Metagenomic sequencing with spiked primer enrichment of general and human-associated fecal indicators for water quality assessment. Water Res. 170, 115748.https://doi.org/10.1016/j.watres.2020.115748.

Anderen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., Garry, R.F., 2020. The proximal origin of SARS-CoV-2. Nat. Med. 26 (4), 450–452.https://doi.org/10.1038/s41591-020-0820-6.

Arras, S., Nag, A., Sethi, J., Rajvanshi, J., Saena, S., Shirvastava, S.K., Gupta, A.B., 2020. Sewage surveillance for the presence of SARS-CoV-2 genome as a useful wastewater based epidemiology (WBE) tracking tool in India. Water Sci. Technol. 82 (12), 2823–2836.https://doi.org/10.2166/wst.2020.540.

Baig, A., Zaiz, A., Ott, I.M., Petrone, M.E., Alpert, T., Fauler, J.R., Kalinich, C.C., Vogels, C.B.F., Breben, M.I., Dunavel, C., McElroy, K., Ghaedi, N., Imakiev, M., Mckenzie-Bennett, M., Robinson, K., Plociak, A., Schilling, R., Pierser, M., Littlefield, R., Baym, M., 2021. Variant Abundance Estimation for SARS-CoV-2 in Wastewater Using RNA-Seq Quantification. https://doi.org/10.1137/21M1426298 (p. 2021.08.31.2126298).

Bloom, J.S., Sathe, L., Munugal, C., Jones, E.M., Gasperini, M., Lubok, N.B., Yarza, F., Thompson, E.M., Kovary, K.M., Park, J., Marquette, D., Kay, S., Lucas, M., Love, T., Sina Bossuyt, A., Brandenborg, O.G., Gao, L., Boeckx, J., Hochman, M., Arboleda-Bar-Or, I., Weil, M., Indenbaum, V., Buciri, E., Bar-Ilan, D., Elad, M., Levi, N., Aguaviva, I., Cohen, Z., Shirazi, R., Erster, O., Sela-Brown, A., Sofer, D., Mor, O., Mendelson, E., Zuckermandl, N.S., 2021. Detection of SARS-CoV-2 variants by genomic analysis of wastewater samples in Israel. Sci. Total Environ. 789, 148002.https://doi.org/10.1016/j.scitotenv.2021.148002.

Brockbank, M.A., Booth, A.B., Salt, M., Spencer, S.K., Wipfipint, K.R., Noble, R.T., 2021. The environmental microbiology minimum information (EMMI) guidelines: qPCR and dPCR quality and reporting for environmental microbiology. Environ. Sci. Technol. 55 (15), 10210–10223.https://doi.org/10.1021/acs.est.1c01767.

Bushnell, B., 2014. BBMap a fast, accurate, splice-alignment (BLNI-7065E), Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).https://www.osti.gov/biblio/1241166.

Cahy, V.D., Torti, S., Yasui, M., Kyowu, S., Katayama, H., 2021. Caspiqgrid RT-qPCR for the selective detection of SARS-CoV-2 in wastewater. Sci. Total Environ. 791, 148342.https://doi.org/10.1016/j.scitotenv.2021.148342.

Cao, Y., Raith, M.F., Griffith, J.F., 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.

Charre, C., Gineva, C., Sabatier, M., Regue, H., Dez, M., Lassen, J., Gros, E., Brun, G., Soltesz, C., Moref, F., Valette, M., Lina, B., Bal, A., Josset, L., 2020. Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterization. Virus Evol. 6 (2), vez075.https://doi.org/10.1093/ve/vez075.

Chiara, M., D’Erchia, A.M., Giovi, C., Manzari, C., Parisi, A., Resta, N., Zambelli, F., Picardi, E., Pavesi, G., Hommer, D.S., Pesole, G., 2021. Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. Brief. Bioinform. 22 (4), 616–630.https://doi.org/10.1093/bib/bbaa297.

Ciesielski, M., Blackwood, D., Clerkin, T., Gonzalez, R., Thompson, H., Larson, A., Noble, R., 2021. Assessing sensitivity and reproducibility of RT-dPCR and RT-qPCR for the quantification of SARS-CoV-2 in wastewater. J. Virol. Methods 297, 114230.https://doi.org/10.1016/j.jviromet.2021.114230.

Coil, D.A., Albertson, T., Banerjee, S., Brennan, G., Campbell, A.J., Cohen, S.H., Dandekar, S., Diaz-Muñoz, S.L., Eisen, J.A., Goldstein, T., Jones, I.R., Juzur, M., Robinson, B.A., Robinson, A.M., Sandrock, R.S., Tompkins, M.M., Tompkins, A., Slack, A., 2021. SARS-CoV-2 detection and genomic sequencing from hospital surface samples collected at UC Davis. PLOS ONE 16 (6), e0253578.https://doi.org/10.1371/journal.pone.0253578.

Cris-Cristophor, A., Kanto, R.S., Olof, M.R., Whitney, O.N., Al-Shayeb, B., Lou, Y.C., Flamholtz, A., Kennedy, L.C., Greenwald, H., Hinkle, A., Hetzal, J., Spitzer, S., Koble, J., Tan, A., Hyde, F., Schroth, G., Kuersten, S., Banfield, J.F., Nelson, K.L., 2021. Genome sequencing of sewage bacteria detects regionally prevalent SARS-CoV-2 variants. Milu (12).https://doi.org/10.1126/mi.002703.20.

D’Aoust, P.M., Mercier, E., Montpetit, D., Jia, J.-J., Alexander, I.D., Neault, N., Baig, A.T., Mayne, J., Zhang, X., Alain, T., Langlois, M.-A., Servos, M.R., Mackenzie, M., Fieges, D., Mackenzie, A.E., Graber, T.E., Delatolla, R., 2021. Quantitative analysis of SARS-CoV-2 RNA from wastewater in communities with low COVID-19 seroprevalence and prevalence. Water Res. 188, 116560.https://doi.org/10.1016/j.watres.2020.116560.

Davis, J.J., Long, S.W., Christensen, P.A., Olson, R.J., Olson, R., Shukla, M., Sedubi, S., Stevens, R., Muster, J.M., 2021. Analysis of the ARTVIC version 5 and version 4 SARS-CoV-2primers and their impact on the detection of the G142D amino acid substitution in the spike protein. Microbiol. Spectrum https://doi.org/10.1128/Spectrum.01803-21.

Diaz-Muñoz, S.L., Achari, A., Federman, S., Yu, G., Somasekar, S., Bártolo, I., Yagi, S., Mbala-Cloppet, A., Cluckow, A., 2021. SARS-CoV-2 detection and genomic sequencing from hospital surface samples collected at UC Davis. PLOS ONE 16 (6), e0253578.https://doi.org/10.1371/journal.pone.0253578.

Díaz-Muñoz, S.L., Eisen, J.A., Gros, E., Valette, M., Lina, B., Bal, A., Josset, L., 2020. Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterization. Virus Evol. 6 (2), vez075.https://doi.org/10.1093/ve/vez075.

Erchia, A.M., Gissi, C., Manzari, C., Parisi, A., Resta, N., Zambelli, F., Picardi, E., Pavesi, G., Hommer, D.S., Pesole, G., 2021. Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. Brief. Bioinform. 22 (4), 616–630.https://doi.org/10.1093/bib/bbaa297.

Ericha, A.M., Giovi, C., Manzari, C., Parisi, A., Resta, N., Zambelli, F., Picardi, E., Pavesi, G., Hommer, D.S., Pesole, G., 2021. Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. Brief. Bioinform. 22 (4), 616–630.https://doi.org/10.1093/bib/bbaa297.

Flood, M.T., D’Souza, N., Rose, J.B., Aw, T.G., 2021. Methods evaluation for rapid concentration and quantification of SARS-CoV-2 in raw wastewater using droplet digital and
