

Estimating relative abundance of two SARS-CoV-2 variants through wastewater surveillance at two large metropolitan sites

Alexander Yu (✉ alexandertyu@gmail.com)

California Department of Public Health <https://orcid.org/0000-0002-5502-9818>

Bridgette Hughes

Verily Life Sciences

Marlene Wolfe

Emory University

Tomas Leon

California Department of Public Health

Dorothea Duong

Verily Life Sciences

Angela Rabe

California Department of Public Health

Lauren Kennedy

Stanford University

Sindhu Ravuri

California Department of Public Health

Bradley White

Verily Life Sciences

Krista Rule Wigginton

University of Michigan

Alexandria Boehm

Stanford University <https://orcid.org/0000-0002-8162-5090>

Duc Vugia

California Department of Public Health

Article

Keywords:

Posted Date: November 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1083575/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Monitoring SARS-CoV-2 variants of concern (VOCs) is critical for public health management of COVID-19. Case isolate sequencing is resource-intensive and not all isolates can be sequenced, and thus are not representative. However, since wastewater SARS-CoV-2 RNA concentrations correlate with COVID-19 case incidence in sewersheds, tracking VOCs through wastewater is appealing. We developed targeted digital RT-PCR assays to monitor abundance of select mutations in Alpha and Delta VOCs in wastewater settled solids, applied these to July 2020-August 2021 samples from two large metropolitan sewersheds, and compared results to estimates of variant abundance from case isolate sequencing. Wastewater measurements tracked closely with case isolate estimates at each site ($r_p = 0.82, 0.88$ for Alpha and $r_p = 0.97$ for Delta). Mutations were detected in wastewater measurements even at levels $<5\%$ of circulating virus and in samples available weeks before case isolate results. Therefore, wastewater variant monitoring should be strategically deployed to complement case isolate sequencing.

Background

The coronavirus disease 2019 (COVID-19) pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has, by September 2021, claimed more than 4.6 million lives worldwide, including $>665,000$ in the United States and $>67,000$ in California¹⁻³. Since its emergence in late 2019, SARS-CoV-2 has mutated genetically resulting in some variants with concerning characteristics that are categorized by the World Health Organization (WHO) as variants of interest (VOI) and of concern (VOC). VOCs have evidence of increased infectiousness and clinical severity, and have spread globally. Some VOCs, such as Alpha and Delta, have, at different times and in different regions, become the predominant strain among cases, hospitalizations, and deaths^{4,5}. There is also concern that current COVID-19 diagnostics, therapeutics, and/or vaccines may have decreased effectiveness against a current or future VOC^{6,7}. By September 2021, VOCs in the U.S. have included the Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), and Delta (B.1.617.2) variants, and since June 2021, Delta has outcompeted Alpha to become the predominant strain circulating in the country³.

Monitoring for SARS-CoV-2 VOCs is a critical component of the response to and management of the ongoing COVID-19 pandemic. This enables public health officials to track their public health impact, implement control measures, and allocate resources effectively. Currently, detection of SARS-CoV-2 viral mutations to identify variants occurs primarily through genomic sequencing of SARS-CoV-2 isolates collected for polymerase chain reaction (PCR)-based diagnosis of people with active COVID-19 infection. Sequencing of these case isolates is resource and time intensive, with limits on sequencing capacity due to equipment, reagents, and trained personnel⁸. As such, complete and timely sequencing of all case isolates is not a feasible or practical goal, particularly when the number of cases has been high. Between January 2020 and September 2021, less than 3% of COVID-19 cases in the US had isolates that were sequenced and available on public repositories³. Non-random selection of isolates for sequencing, including due to geographically disparate sequence infrastructure and preferential allocation of

sequencing resources for clinical and public health priorities (e.g., associated with an outbreak or with severe illness), and non-uniform result reporting, could make sequencing results susceptible to bias and not truly representative of circulating variants within communities^{4,8,9}. In addition, there can be substantial time delay between when an isolate is collected and sequenced and when results become available for public health response⁹. Due to its timeliness and representativeness, with comparatively low costs, wastewater surveillance for SARS-CoV-2 VOCs can be a useful supplement to sequencing surveillance based on case isolates¹⁰⁻¹².

Since early in the pandemic, wastewater samples have been collected and analyzed to quantify the amount of SARS-CoV-2 viral RNA in sewage, with estimates of viral abundance in the sewage catchment area (sewershed) correlating closely with reported COVID-19 case counts for the same area^{13,14}. Wastewater surveillance can provide a more complete snapshot of real-time community transmission of viral activity independent of individual care seeking or testing behavior. Given that, there is a strong interest in determining if wastewater can also provide useful information on circulating VOCs¹⁵. Both sequencing and use of PCR assays targeting specific mutation(s) have been proposed as methods to detect mutations and deletions in SARS-CoV-2 RNA in wastewater.

However, variant monitoring using these environmental samples present technical challenges. Variants are characterized by the presence of multiple mutations or deletions on the same RNA genome, and some variants share one or more mutations¹⁶. Unlike SARS-CoV-2 isolates from an individual case which consist of a single genome, wastewater samples likely contain composite material from multiple circulating SARS-CoV-2 variants shed from different individuals, with each genome at low concentration and in various states of genomic integrity due to degradation in wastewater¹⁷. This results in decreased sequence coverage breadth when using next generation and amplicon sequencing approaches, lowering confidence in detection of variants and generation of data that is compositional in structure and not directly quantitative. In contrast, targeted PCR approaches have the potential to detect mutations present in variants at low concentrations (<10 copies per reaction)¹⁸. However, since wastewater contains a complex mixture of SARS-CoV-2 RNA fragments, the presence of one (or more) mutation sequences present within the genome of a variant does not alone prove that the variant is present in wastewater. Currently, the presence of variants in wastewater must be inferred. Consequently, the interpretation of variant mutation data to estimate variant relative abundance in wastewater using targeted assays and sequencing approaches is an ongoing area of research.

As a proof of concept, we developed targeted digital RT-PCR mutation assays to retrospectively and prospectively monitor for the presence and abundance of mutations present in the Alpha (B.1.1.7) and Delta (B.1.617.2) VOCs in wastewater settled solids. Wastewater solids are chosen for this work as they contain orders of magnitude higher concentrations of viral RNA than wastewater influent^{19,20}, and previous work has documented a strong coupling between concentrations of SARS-CoV-2 RNA in wastewater solids and incidence rates in the population contributing to the wastewater²⁰. We prospectively monitored wastewater solids of a large metropolitan sewershed in California between July

2020 and August 2021 for a deletion present in the Alpha variant. We then retrospectively measured the abundance of this deletion in a second large metropolitan area where samples had been routinely collected, and also measured concentrations of mutations suggestive of Delta in both sewersheds. This data was then compared against estimates of Alpha and Delta abundance in each of these sewersheds derived from COVID-19 case isolate sequencing data available to the California Department of Public Health (CDPH).

Methods

HV69-70 and Del156-157/R158G mutation assay development for Alpha and Delta variants. The HV69-70 and Del156-157/R158G mutation assays were developed in silico using Primer3Plus (<https://primer3plus.com/>) to target mutations present in alpha and delta variants of concern, respectively. These mutations were chosen because they are present in high percentages of the associated variant sequences in GISAID, and they are deletions; PCR assays targeting deletions are generally more specific than those targeting single nucleotide polymorphisms (SNPs). Mutation sequences were obtained from outbreak.info/compare-lineages. The parameters used in the development (that controlled sequence length, GC content, and melt temperatures) are provided in the Supplementary Information (SI). Primers and probe sequences are provided in Table 1 and schematic of the assays are provided in Figure S1, S2, and S3.

Primers and probe sequences were screened for specificity in silico using NCBI Blast, and then tested in vitro against a virus panel (NATtrol™ Respiratory Verification Panel, Zeptomatrix) that includes several influenza and coronavirus viruses, “wild-type” gRNA from SARS-CoV-2 strain 2019-nCoV/USA-WA1/2020 (ATCC® VR-1986D™) which does not contain the mutations (hereafter referred to as WT-gRNA), heat inactivated SARS-CoV-2 strain Alpha (SARS-CoV-2 variant B.1.1.7, ATCC® VR-3326HK™), and synthetic gRNA from Twist Biosciences (South San Francisco, California, USA) for Beta (Twist control 16), Gamma (Twist control 17), Kappa (Twist control 18), and Delta (Twist control 23) variants. Details on sample processing are provided in the SI. The sensitivity and specificity of the mutation assays were further tested by diluting variant gRNA containing the mutations in no (0 copies), low (100 copies), and high (10,000 copies) background of WT-gRNA (details in SI).

Wastewater sample collection. Two publicly owned treatment works (POTWs) that serve populations of Santa Clara County, California, USA (San Jose, SJ) and Sacramento County, California, USA (Sacramento, SAC) were included in the study. Each serves approximately 1,500,000 residents; further descriptions of these POTWs can be found in Wolfe et al.¹⁴.

Samples were collected by POTW staff using sterile technique in clean, labeled bottles. Approximately 50 ml of settled solids were collected each study day from each POTW. Settled solids samples were grab samples at SAC. At SJ, POTW staff manually collected a 24 h composite sample. Samples were immediately stored at 4°C and transported to the lab and processed (within 6 hours of collection).

Samples were collected daily for a larger wastewater surveillance effort starting in October 2020¹⁴, and a subset of these samples are used in the present study. The dates of samples included in this study were chosen to span the period of time prior to and including the presumed emergence of Alpha and Delta variants in the communities. Prior to presumed emergence of Alpha and Delta variants, 1-4 samples per month were analyzed, and during periods thought to correlate with emergence and spread, sampling was daily to three times per week. At SJ, 133 and 48 samples were included for HV69-70 and del156-157/R158G analyses, respectively. At SAC, 64 and 48 samples were included for HV69-70 and del156-157/R158G analyses, respectively. Samples from as early as July 2020 and as late as August 2021 were included.

Samples were processed as described in Wolfe et al.²¹. Briefly, solids were dewatered by centrifugation and a dry weight measurement taken. Solids were resuspended in a DNA/RNA shield solution (Zymo Research) spiked with Bovine Coronavirus (BCoV) as a positive recovery control and homogenized. Prior to 26 May 2021, RNA was extracted from the homogenized sample. Thereafter, the final vortexing step was omitted and RNA was extracted from the resultant supernatant. Comparative analysis suggests that removing the final vortexing step reduced variability among replicates. RNA was extracted using the Chemagic™ Viral DNA/RNA 300 Kit H96 for the Perkin Elmer Chemagic 360 followed by PCR Inhibitor Removal with the Zymo OneStep-96 PCR Inhibitor Removal Kit. RNA was subsequently processed immediately (within 24h of sample collection) to measure concentrations of the N gene, PMMoV, and BCoV recovery. The N gene codes for the SARS-CoV-2 nucleocapsid and the specific region of the genome targeted by our assay is conserved on all SARS-CoV-2 genomes. PMMoV is highly abundant in human stool and domestic wastewater globally^{22,23} and is used here as an internal recovery and fecal strength control for the wastewater samples. Subsequently, RNA samples from San Jose (SJ) were stored between 0 and 7 days before they were run for HV69-70 and the N gene in a multiplex assay. RNA samples from SJ and Sacramento were stored between 15 and 190 days prior to being analyzed for del156-157/R158G and the N gene; and RNA samples from Sacramento were stored between 15 and 300 days prior to being analyzed for HV69-70 and the N gene. The N gene was run a second time in all samples as a check for RNA degradation during storage.

Further details are available in the SI.

ddRT-PCR. Digital droplet RT-PCR assays for the mutations and N gene targets (see Table 1 for primer and probe sequences, purchased from IDT) were performed on 20 µl samples from a 22 µl reaction volume, prepared using 5.5 µl template, mixed with 5.5 µl of One-Step RT-ddPCR Advanced Kit for Probes (Bio-Rad 1863021), 2.2 µl Reverse Transcriptase, 1.1 µl DTT and primers and probes at a final concentration of 900 nM and 250 nM respectively. Droplets were generated using the AutoDG Automated Droplet Generator (Bio-Rad). PCR was performed using Mastercycler Pro with cycling conditions described in the SI. Droplets were analyzed using the QX200 Droplet Reader (Bio-Rad). All liquid transfers were performed using the Agilent Bravo (Agilent Technologies).

Positive controls consisted of BCoV and PMMoV gene block controls (dsDNA purchased from IDT), gRNA of SARS-CoV-2 (strain 2019-nCoV/USA-WA1/2020, ATCC® VR-1986D™), gRNA of SARS-CoV-2 Alpha lineage (SARS-CoV-2 variant B.1.1.7, ATCC® VR-3326HK™), and gene block controls for Del156-157/R158G. Details of the number of wells assayed for samples and controls are listed in the SI. Results from replicate wells were merged for analysis. Thresholding was done using QuantaSoft™ Analysis Pro Software (Bio-Rad, version 1.0.596). In order for a sample to be recorded as positive, it had to have at least 3 positive droplets. For the wastewater samples, three positive droplets corresponds to a concentration between ~500-1000 cp/g; the range in values is a result of the range in the equivalent mass of dry solids added to the wells.

For the wastewater samples, concentrations of RNA targets were converted to concentrations per dry weight of solids in units of copies/g dry weight using dimensional analysis. The total error is reported as standard deviations and includes the errors associated with the Poisson distribution and the variability among the 10 replicates. The recovery of BCoV was determined by normalizing the concentration of BCoV by the expected concentration given the value measured in the spiked DNA/RNA shield. BCoV was used solely as a process control; samples were rerun in cases where the recovery of BCoV was less than 10%. All wastewater data, along with additional reporting details per the MIQE²⁴ and EMMI²⁵ guidelines are available publicly at the Stanford Digital Repository (<https://doi.org/10.25740/zf117dn1545>).

Incident COVID-19 cases and COVID-19 case isolate sequences. Sewershed service area boundary shapefiles were provided by each POTW. Sewershed boundaries were compared to zip code boundaries (2020 US Census TIGER/Line shapefiles) in ArcGIS Pro version 2.7.3 (ESRI, Redlands CA). All ZIP codes with >50% of land area within sewershed boundaries were considered to be within the POTW service area and used for further analysis. The number of PCR-confirmed COVID-19 cases reported to CDPH as a function of episode date (earliest of either specimen collection date or date of symptom onset) residing within each sewershed was determined using methods reported previously²¹.

COVID-19 case isolate whole genome sequence data used in this study were sourced from results provided to CDPH by the CDC, which included results through laboratory partners contracted to sequence case isolates across the United States. Sequence data were assigned to a sewershed based on ZIP code provided for the sample, with the time of sample collection assigned as date. The PANGO lineage was assigned based on the version available at the time data was extracted, with the most recent results using pangoleARN and pangole-designation v1.2.66²⁶.

Estimates of VOC abundance between February 1 and July 31, 2021 were calculated by dividing the number of sequences identified as Alpha or Delta (according to WHO definition and including all PANGO sublineages Q.* and AY.*, respectively) by the total number of isolates sequenced from individuals residing in the sewersheds with specimens collected over the previous 14 days. To estimate the proportion of total COVID-19 cases with an isolate that was sequenced, we divided the total number of isolates sequenced by the number of COVID-19 cases reported to CDPH, by episode date, over the same 14-day period and sewershed. To estimate the time needed between isolate sample collection,

sequencing, and report receipt, and the impact of that time delay on VOC estimates for the sewershed community, we compared changing 14-day VOC abundance estimates over time against a final estimate that was generated after the end of the analysis period (August 24, 2021).

Pearson correlations were performed between the wastewater mutation and case isolate variant data sets, comparing the mean ratio of mutation/deletion(s) in wastewater (HV69-70 and del156-157/R158G to the N gene) to the proportion of case isolates sequenced that were characterized as alpha or delta, each averaged over the previous 14 days. Statistical significance was set at $p < 0.05$; analyses were carried out using R studio Version 1.4.1106.

Results

Variant mutation assay specificity and sensitivity. In silico analysis indicated no cross reactivity between the assays and deposited sequences in NCBI. When challenged against WT-gRNA, the respiratory virus panel, and actual or synthetic variant gRNA, no cross reactivity was observed. Positive controls and no template controls run on the sample plate performed according to expectations. Variant mutation concentrations were measured in no, low, and high background of WT-gRNA that does not contain the mutations. Results of mutation assays in the presence of high and low background WT-gRNA are similar to their results in the absence of background WT-gRNA (Figure SX). We were able to detect concentrations as low as 5 copies/well for the mutation assays when there was no, low, and high background of WT-gRNA. These results together suggest that the mutation assays are specific and sensitive, even in the presence of high levels of background SARS-CoV-2 RNA.

Variant mutation concentrations in wastewater solids. Positive and negative controls were as expected, and recovery of BCoV was higher than 10% and PMMoV within expected ranges previously measured at the plant indicating consistent recovery of RNA from samples and lack of substantial inhibition. Additional details of controls are provided in the SI.

HV69-70 concentrations were measured up to daily at SJ and up to three times per week at SAC; concentrations ranged from not detected to over 10,000 copies/g (Figure 1 and 2). N and HV69-70 concentrations at SJ prior to 15 Feb 2021 are not presented graphically; samples collected during July and September 2020 did not have measurable HV69-70. HV69-70 was measured for the first time in SJ solids in November 2020 at concentrations $\sim 10^3$ copy/g. HV69-70 was not detected in SAC wastewater solids prior to late February 2021; samples collected in October 2020 (not shown in top plot) and late January 2021 were non-detect for HV69-70. At both locations, the concentration of HV69-70 relative to the N gene (HV69-70/N ratio) increased over time beginning in early March 2021, peaked in early June 2021 at SJ and May 2021 in SAC when HV69-70/N was ~ 1 , and then fell until HV69-70 became undetectable at SJ and present at very low relative concentrations at SAC (0.01) in late July 2021 (Figures 1 and 2).

Del156-157/R158G mutation concentrations were measured as frequently as 3 times per week at both SJ and SAC and ranged from not detected to 100,000 copies/g (Figures 1 and 2). Del156-157/R158G non-detects were observed in samples collected prior to early April 2021 at both sites, and then both sites experienced a small peak in Del156-157/R158G concentration in early to mid-May 2021, followed by a decline to non-detect, followed by a sharp increase until the end of the data series. During this time, N gene concentrations in wastewater increased contemporaneously. The concentration of del156-157/R158G relative to N (del156-157/R158G/N ratio, Figures 1 and 2) increased to approximately 0.8 at the sites by the end of the data series.

Trends in presence of variants in sequenced case isolates from sewersheds. The trends in confirmed presence of Alpha and Delta variants in isolates collected from residents (case isolates) of the SJ and SAC sewersheds between early February and late July 2021 are provided in Figure 3. Alpha proportions increased in both sewersheds from early March, peaking in May/June and decreasing in early July. Delta was first identified in isolates in early April and by the end of July, accounted for almost all sequenced isolates. In SJ, a small peak in Delta was observed in May, prior to its large, sustained increase in June; a similar peak is also evident, to a lesser extent, in the SAC Delta data. During this period of time, the 7-day average laboratory-confirmed incident COVID-19 cases ranged from 1 to 30 per 100,000 (Figure SD) in each sewershed. Incident COVID-19 cases in each sewershed is positively and significantly correlated with N gene measurements in the settled solids ($r_p = 0.8$, $p < 10^{-10}$ for both SJ and SAC N gene data sets, regardless of whether they were generated when measuring the delta or alpha mutation).

Relationship between proportion of Alpha and Delta variants in case isolates and wastewater mutation data. Ratios of HV69-70 and del156-157/R158G mutations to the N gene (VOC abundance estimates based on wastewater) were compared against the proportion of all case isolates sequenced that were identified as Alpha and Delta variant (VOC abundance estimates based on case isolate sequencing) from each sewershed, respectively (Figure 3). The trends of the wastewater mutation ratios follow closely and temporally the trends of the proportion of Alpha and Delta variants during this period at both sewersheds, including features such as an early peak in Delta in May. Alpha and Delta mutation gene ratios from wastewater were strongly correlated to the corresponding ratios of each VOC from case isolates sequenced: $r_p = 0.82$ ($p < 10^{-5}$, $df = 19$, SJ) and 0.88 ($p < 10^{-7}$, $df = 21$, SAC) for Alpha; $r_p = 0.97$ ($p < 10^{-15}$, $df = 23$, for both SJ and SAC) for Delta. When compared to the opposing variant, the mutation gene ratios were not correlated ($p > 0.05$ for all).

Completeness of and delays in receiving SARS-CoV-2 isolate sequence data. From February 1 - August 1, 2021, the total number of case isolates that were sequenced over a 14-day period in our two sewersheds varied from 2 to 520, with a median estimate of 8% of all case isolates in SJ and 6% in SAC sequenced. Earliest isolate sequencing results were available to CDPH around five days after sample collection date. Approximately 75% of all sequenced isolate results in our dataset were available within 2-3 weeks. As more isolate sequencing data was received for each 14-day period, estimated proportions of VOCs changed over time. Around three weeks were required for 95% of 14-day window VOC estimates to be within 10% of the final estimate.

Discussion

Our results show that the HV69-70 and del156-157/R158G mutation assays as used for wastewater settled solids were sensitive and specific, and we found strong correlation between wastewater estimates of circulating Alpha and Delta VOCs using these PCR mutation assays with estimates of these variants from case isolate sequencing data in two large metropolitan communities in California. When the mutation assays were used, prospectively for Alpha and retrospectively for Delta wastewater samples, mutations were detected in wastewater samples collected 1-3 weeks earlier than when Alpha and Delta variant estimates generated by case-isolate sequencing were available and reliable. Targeted mutation assays applied to SARS-CoV-2 RNA extracted from wastewater solids can be a rapid, efficient, and reliable way to monitor COVID-19 VOCs introduced to and circulating in a community. Monitoring for those VOCs using wastewater may provide earlier complementary surveillance data than from case isolate sequencing data, if mutation assays for new VOCs can be developed and implemented in a timely manner.

Use of PCR assays targeting characteristic mutation/deletion(s) thought to be particular to a SARS-CoV-2 variant may concurrently detect other SARS-CoV-2 strains that carry the same mutation/deletion(s). Targeting a single mutation in wastewater, as was done here for Alpha, carries an increased potential risk for mischaracterization. For example, on Sept 8, 2021, according to GISAID, a global repository of case isolate sequence data, 1,043,561 (97%) of the 1,077,360 Alpha (B.1.1.7 and Q sublineages) sequences from COVID-19 case isolates contained the HV69-70 mutation. However, HV69-70 was also present in other variants such as B.1.258.19 where it was present in all 141 B.1.258.19 sequences in GISAID, and B.1.617.2 where it was present in 647 (0.2%) of 402,038 sequences. Targeting multiple mutations, as was done here with Delta, can increase specificity. Of the 937,570 sequences in GISAID classified as Delta (B.1.617.2 and AY sublineages), 842,354 (90%) have the del156-157/R158G mutations (referred to as E156G/del157-158 in GISAID). While this combination of mutations can also be present in other variants, it is rarer; the non-delta variant with the highest percentage of sequences with these mutations is B.1.617.3, for which there were 266 isolates in the global GISAID database and only 77 (29%) of them possessing these mutations. The non-delta variant with these mutations for which there are the largest number of isolates in GISAID is B.1.1.7 for which 6 (0.0006%) of the over 1,053,637 million sequences have these mutations.

Our study findings show that use of the HV69-70 and del156-157/R158G mutation assays to estimate for circulating Alpha and Delta variants in wastewater, respectively, has correlated well with estimates of these VOCs from case isolate sequencing data. Wastewater estimates for Alpha, based on a single deletion assay, were robust over time in two large municipalities over eight months ($r_p = 0.82$, $p < 10^{-5}$ in SJ and 0.88 , $p < 10^{-7}$ in SAC, including periods of high (tail of 2020 winter and 2021 summer) and low (2021 spring) community SARS-CoV-2 transmission. Similarly, estimates for Delta, based on multiple mutation/deletion(s), correlated highly with estimates from sequenced case isolates ($r_p = 0.97$, $p < 10^{-15}$, for both SJ and SAC). Concurrent monitoring of VOCs in both wastewater and in case isolates can confirm

whether the targeted mutation assays used are correlated with the VOCs being monitored and mitigate risks of misinterpreting wastewater results. Discrepant or divergent estimates between the two datasets should be noticeable within weeks and would then suggest another variant with the same mutations circulating at abundance, prompting investigation if unexpected.

For validated assays deployed in established wastewater sites, wastewater surveillance for SARS-CoV-2 VOCs could be an important adjunctive estimate of variant circulation to add to estimates derived from case isolate sequencing data. Because cost and limited genomic testing capacity make sequencing all COVID-19 isolates impractical, especially during times of high case incidence, health departments and decision makers extrapolate variant information from relatively small numbers or proportions of sequenced isolates, which may be biased and unrepresentative. For our case dataset, 14-day VOC estimates were derived from as few as 2-20 total case isolates and <1% of estimated total case isolates sequenced.

Wastewater variant monitoring can overcome biases and delays seen with case isolate sequencing. Since everyone living in a sewershed contributes waste to the system, wastewater monitoring is independent of testing and care accessibility biases and results are more representative of COVID-19 cases in that sewershed. In addition, mutation assay results are available from collected wastewater in a shorter time than VOC estimates from sequencing of case isolates. In our monitored sewersheds, the total average turnaround time from collection of wastewater to testing results was less than 8 hours. In contrast, for our two sewersheds, it took 2-3 weeks after sample collection date for 75% of case isolate sequence results to be received and three weeks for most 14-day VOC estimates to be within 10% of their final estimate. This does not include additional time delay between case symptom onset and test taking that could further accentuate time advantages of wastewater variant monitoring.

There are important limitations to use of wastewater for SARS-CoV-2 variant monitoring. Laboratory limits of detection for SARS-CoV-2 RNA in wastewater as well as for targeted mutation/deletion(s) may result in no detection, especially at times when low COVID-19 case counts in the population mean lower overall concentrations of SARS-CoV-2 RNA in wastewater. However, even in mid-May 2021, when case counts in our sewersheds were as low as 1-2 per 100,000 people, both SARS-CoV-2 RNA levels and variant abundance could still be measured and accurately estimated. Estimates of circulating Alpha and Delta mutations were also able to be consistently detected even at levels <5%. Limits of detection, both of SARS-CoV-2 and for different mutations associated with variants, are likely to vary depending on laboratory methods used as well as which mutation is targeted; delineation of these limits for each laboratory, sewershed, and assay is important for interpreting what a non-detect implies about variant circulation.

Because newly identified VOCs and VOIs to be monitored require new mutation assays to be designed, the time needed to design, test and implement assays is also an important consideration in decisions to monitor for variants in wastewater (Figure SX). While the time to design an assay in silico (< 1 day) and test its sensitivity and specificity in vitro (3-5 days) is short, the time to receive reagents including

synthesized oligos and positive control RNA from vendors can take 4-6 weeks due to supply chain issues and increased demand during the pandemic. Additionally, before an assay can be designed, variant sequences and mutations must be accurately characterized which can delay the start of the assay design process. Efforts to develop assays before variants become VOCs and proactively order reagents can help ensure assays are available for wastewater testing to aid the public health response to the pandemic.

Monitoring for COVID-19 VOCs will continue to be an important public health function in the foreseeable future, a need that will become more salient if SARS-CoV-2 testing of cases and sequencing resources or utilization decrease over time. Difficulty in surveillance based on case isolate sequencing, including due to non-representative sampling and delayed results, mean that additional complementary variant surveillance methods are important. Detection and monitoring of variants in wastewater has been proposed as an adjunct methodology, and our experiences monitoring for two VOCs in two large California municipalities support the use of targeted PCR mutation assays as a useful method to estimate abundance of circulating VOCs and inform public health. In conjunction with continued COVID-19 case isolate sequencing, wastewater variant monitoring can be strategically deployed as an adjunct public health surveillance tool.

Declarations

Competing Interests

Bradley White, Dorothea Duong and Bridgette Hughes are employees of Verily Life Sciences.

Acknowledgements

This work is supported in part by a gift from the CDC-Foundation. Numerous people contributed to wastewater sample collection including Srividhya Ramamoorthy (Sac), Michael Cook (Sac), Ursula Bigler (Sac), James Noss (Sac), Lisa C. Thompson (Sac), Payak Sarkar (SJ), Noel un Enoki (SJ), and Amy Wong (SJ). We also would like to thank the many groups involved in the sequencing of case isolates, including California State, Santa Clara and Sacramento County Public Health departments and the many public, private, and academic laboratories who have dedicated substantial resources towards the understanding of SARS-CoV-2 variants. In addition, we thank the California Department of Public Health COVID-19 Epidemiology and Data teams for their help with COVID-19 data. Parts of this study were performed at Stanford University, which sits on the ancestral and unceded lands of the Muwekma Ohlone people. We pay our respects to them and their Elders, past and present, and are grateful for the opportunity to live and work here.

References

1. California, S. of. Tracking COVID-19 in California. <https://covid19.ca.gov/state-dashboard/>.
2. WHO Coronavirus (COVID-19) Dashboard. <https://covid19.who.int>.

3. CDC. COVID Data Tracker. *Centers for Disease Control and Prevention* <https://covid.cdc.gov/covid-data-tracker> (2020).
4. Campbell, F. *et al.* Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Eurosurveillance* **26**, 2100509 (2021).
5. Twohig, K. A. *et al.* Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. *The Lancet Infectious Diseases* (2021) doi:10.1016/S1473-3099(21)00475-8.
6. Sanders, R. W. & Jong, M. D. de. Pandemic moves and countermoves: vaccines and viral variants. *The Lancet* **397**, 1326–1327 (2021).
7. World Health Organization. Regional Office for Africa. Weekly Bulletin on Outbreak and other Emergencies: Week 39: 21 - 27 September 2020. (2020).
8. Walensky, R. P., Walke, H. T. & Fauci, A. S. SARS-CoV-2 Variants of Concern in the United States—Challenges and Opportunities. *JAMA* **325**, 1037–1038 (2021).
9. World Health Organization. Regional Office for Europe. *Methods for the detection and identification of SARS-CoV-2 variants, March 2021*. <https://apps.who.int/iris/handle/10665/340067> (2021).
10. Crits-Christoph, A. *et al.* Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants. *mBio* **12**, e02703-20.
11. Izquierdo-Lara, R. *et al.* Monitoring SARS-CoV-2 Circulation and Diversity through Community Wastewater Sequencing, the Netherlands and Belgium. *Emerg Infect Dis* **27**, 1405–1415 (2021).
12. Bar-Or, I. *et al.* Detection of SARS-CoV-2 variants by genomic analysis of wastewater samples in Israel. *Sci Total Environ* **789**, 148002 (2021).
13. Kirby, A. E. Using Wastewater Surveillance Data to Support the COVID-19 Response – United States, 2020–2021. *MMWR Morb Mortal Wkly Rep* **70**, (2021).
14. Wolfe, M. K. *et al.* Scaling of SARS-CoV-2 RNA in Settled Solids from Multiple Wastewater Treatment Plants to Compare Incidence Rates of Laboratory-Confirmed COVID-19 in Their Sewersheds. *Environ. Sci. Technol. Lett.* **8**, 398–404 (2021).
15. Peccia, J. *et al.* Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nat Biotechnol* **38**, 1164–1167 (2020).
16. Chakraborty, C., Bhattacharya, M. & Sharma, A. R. Present variants of concern and variants of interest of severe acute respiratory syndrome coronavirus 2: Their significant mutations in S-glycoprotein, infectivity, re-infectivity, immune escape and vaccines activity. *Reviews in Medical Virology* **n/a**, e2270.
17. Wurtzer, S. *et al.* Several forms of SARS-CoV-2 RNA can be detected in wastewaters: Implication for wastewater-based epidemiology and risk assessment. *Water Research* **198**, 117183 (2021).
18. Lee, W. L. *et al.* Quantitative SARS-CoV-2 Alpha Variant B.1.1.7 Tracking in Wastewater by Allele-Specific RT-qPCR. *Environ. Sci. Technol. Lett.* **8**, 675–682 (2021).

19. Li, B., Di, D. Y. W., Saingam, P., Jeon, M. K. & Yan, T. Fine-Scale Temporal Dynamics of SARS-CoV-2 RNA Abundance in Wastewater during A COVID-19 Lockdown. *Water Research* **197**, 117093 (2021).
20. Graham, K. E. *et al.* SARS-CoV-2 RNA in Wastewater Settled Solids Is Associated with COVID-19 Cases in a Large Urban Sewershed. *Environ. Sci. Technol.* **55**, 488–498 (2021).
21. Wolfe, M. K. *et al.* High-Frequency, High-Throughput Quantification of SARS-CoV-2 RNA in Wastewater Settled Solids at Eight Publicly Owned Treatment Works in Northern California Shows Strong Association with COVID-19 Incidence. *mSystems* **0**, e00829-21.
22. Colson, P. *et al.* Pepper Mild Mottle Virus, a Plant Virus Associated with Specific Immune Responses, Fever, Abdominal Pains, and Pruritus in Humans. *PLOS ONE* **5**, e10041 (2010).
23. Rosario, K., Symonds, E. M., Sinigalliano, C., Stewart, J. & Breitbart, M. Pepper Mild Mottle Virus as an Indicator of Fecal Pollution. *Applied and Environmental Microbiology* **75**, 7261–7267 (2009).
24. dMIQE Group & Huggett, J. F. The Digital MIQE Guidelines Update: Minimum Information for Publication of Quantitative Digital PCR Experiments for 2020. *Clin Chem* **66**, 1012–1029 (2020).
25. Borchardt, M. A. *et al.* The Environmental Microbiology Minimum Information (EMMI) Guidelines: qPCR and dPCR Quality and Reporting for Environmental Microbiology. *Environ. Sci. Technol.* **55**, 10210–10223 (2021).
26. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* **5**, 1403–1407 (2020).

Tables

Target	Primer/Probe	Sequence
N Gene	Forward	CATTACGTTTGGTGGACCCT
	Reverse	CCTTGCCATGTTGAGTGAGA
	Probe	CGCGATCAAACAACGTCGG (5' FAM/ZEN/3' IBFQ)
BCoV	Forward	CTGGAAGTTGGTGGAGTT
	Reverse	ATTATCGGCCTAACATACATC
	Probe	CCTTCATATCTATACACATCAAGTTGTT (5' FAM/ZEN/3' IBFQ)
PMMoV	Forward	GAGTGGTTTGACCTTAACGTTTGA
	Reverse	TTGTCGGTTGCAATGCAAGT
	Probe	CCTACCGAAGCAAATG (5' HEX/ZEN/3' IBFQ)
HV69-70 del (Alpha)	Forward	ACTCAGGACTTGTTCTTACCT
	Reverse	TGGTAGGACAGGGTTATCAAAC
	Probe	ATGCTATCTCTGGGACCAAT (5' FAM or HEX/ZEN/3' IBFQ)
del156- 157/R158G (Delta)	Forward	ATTCGAAGACCCAGTCCCTA
	Reverse	AGGTCCATAAGAAAAGGCTGA
	Probe	TGGATGGAAAGTGGAGTTTATTCTAG (5' FAM or HEX/ZEN/3' IBFQ)

Table 1. Primer and probe sequences.

Figures

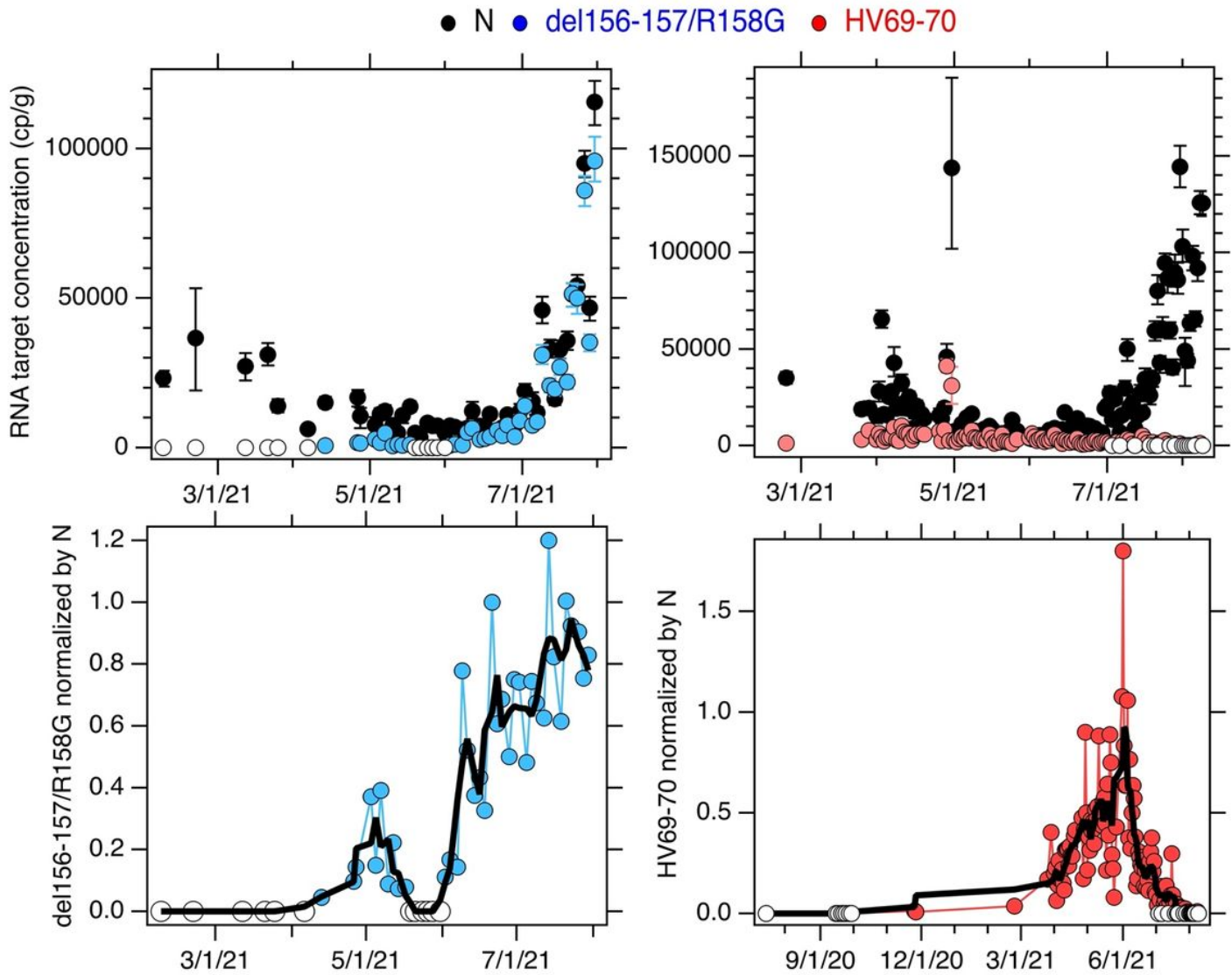


Figure 1

Measurements in SJ wastewater solids. Concentrations of N gene and mutations found in Delta (del156-157/R158G), and Alpha (HV69-70) in wastewater solids (top panels), and their ratio (bottom panels). Errors in top panels represent standard deviations derived from the 10 replicates run for each sample, open white circles are non-detects and shown as 0. Errors include technical and replication errors. If error bars are not visible, then errors are smaller than the symbol. Bottom plots show smoothed lines for visual reference for mutation ratios. For del156-157/R158G/N ratio, the smoothed line is a 3 point running average, and for the HV69-70 / N ratio, the smoothed line is a 7 point running average; each approximates a weekly average. The time scale for the HV69-70 data in the top panel is truncated to facilitate visualization, additional data on dates prior to 2/15/21 are described in the text and shown in the bottom panel.

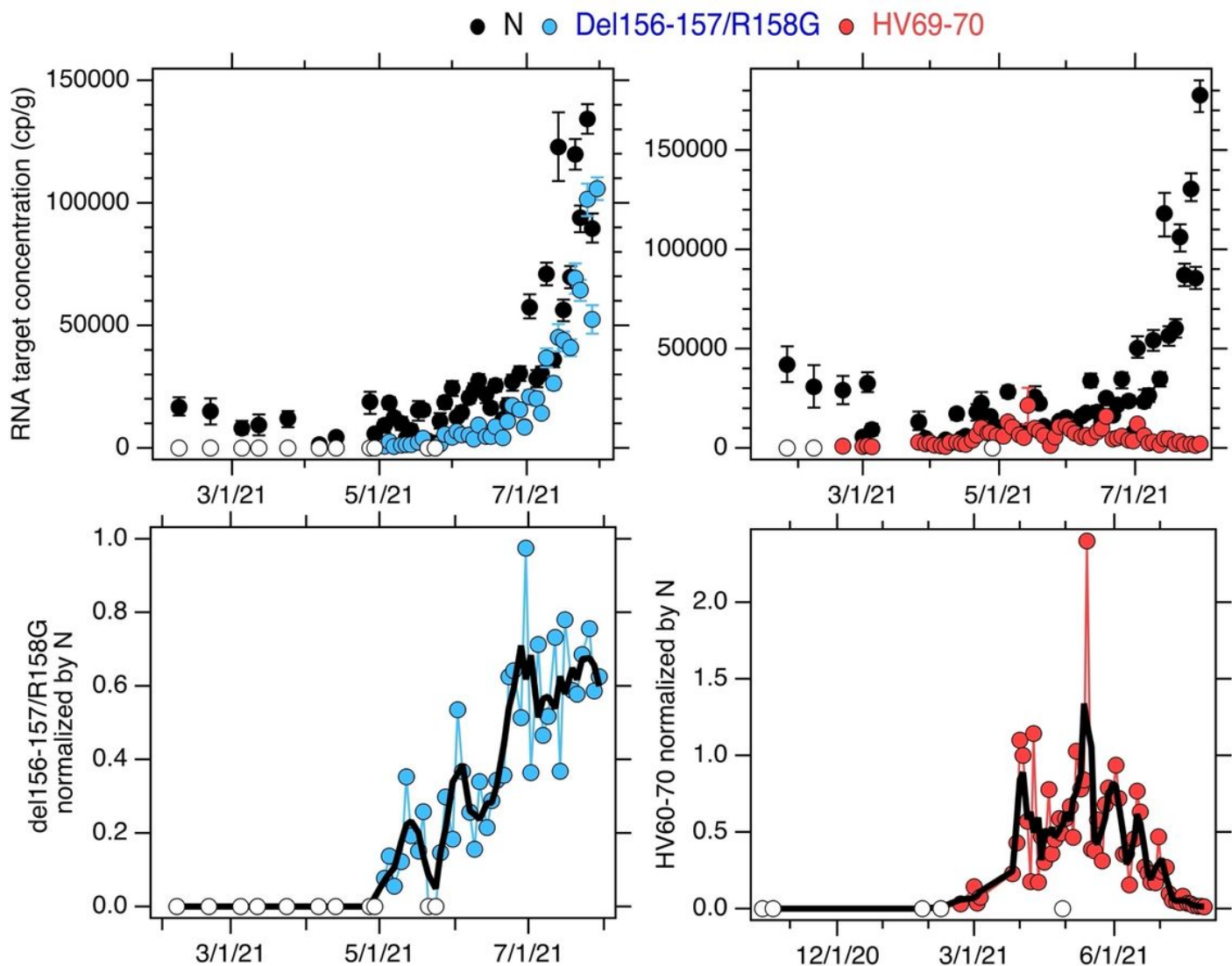


Figure 2

Measurements in SAC wastewater solids. Concentrations of N and mutations found in Delta (del156-157/R158G), and Alpha (HV69-70) in wastewater solids (top panels), and their ratio (bottom panels). Errors in top panels represent standard deviations derived from the 10 replicates run for each sample, open white circles are non-detects and shown as 0. Errors include technical and replication errors. If error bars are not visible, then errors are smaller than the symbol. Bottom plots show smoothed lines (3 point, centered, running average) to provide an approximately weekly average. The time scale for the HV69-70 data in the top panel is truncated to facilitate visualization, additional data on dates prior to 1/15/21 are described in the text and shown in the bottom panel.

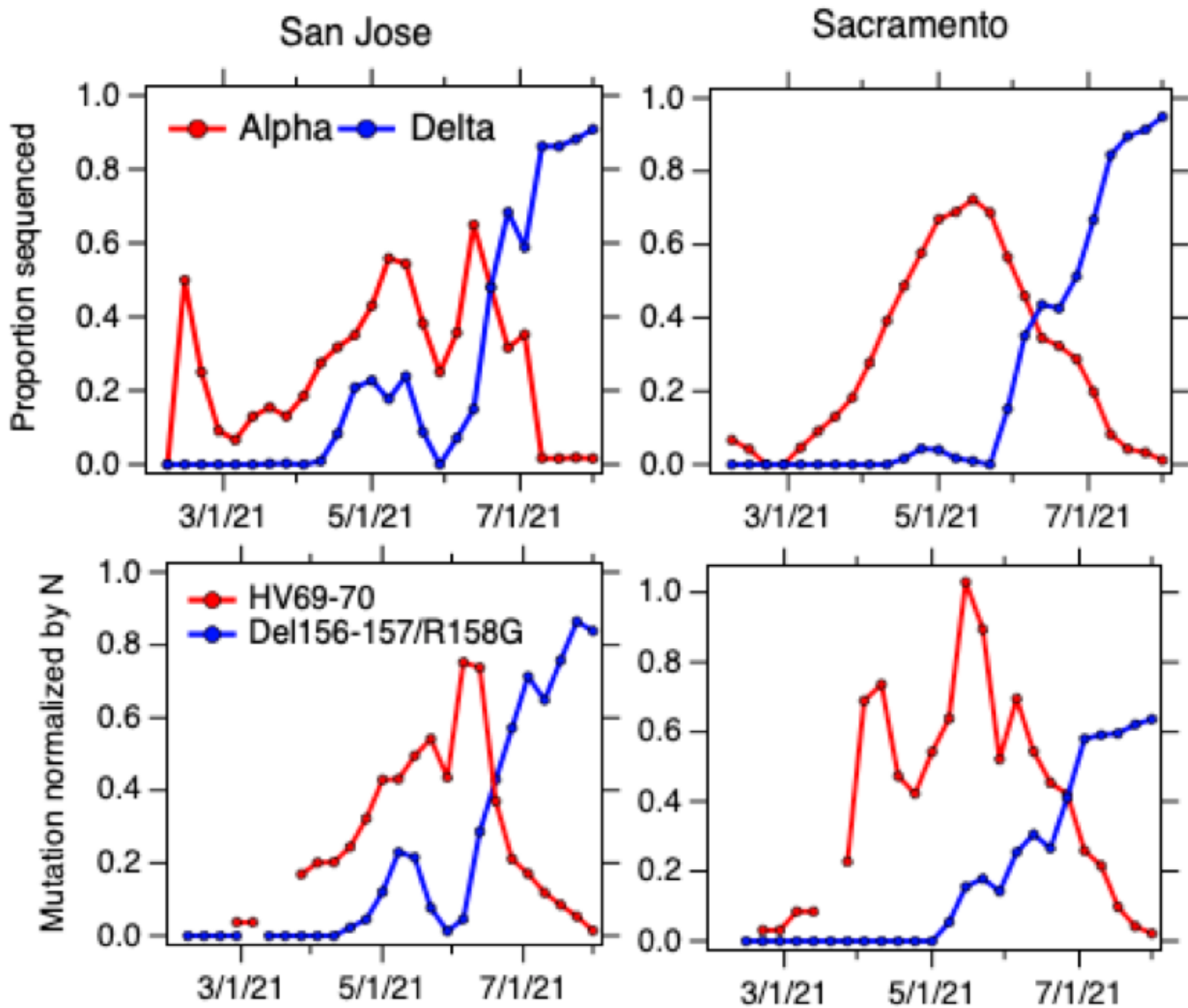


Figure 3

Comparison of Variants of Concern in wastewater with COVID-19 case isolates. Proportion of circulating SARS-CoV-2 virus attributable to the Alpha (red) and Delta (blue) variants, estimated from isolate sequencing data from cases in sewerheds (top row), and estimates using variant mutations in wastewater (bottom row) between Feb 1, 2021 and July 31, 2021. Top row: Proportion of all case isolates collected and sequenced over the previous 14-day interval that are classified as Alpha and Delta in the SJ and SAC sewerheds. Bottom row: Concentrations of mutations found in Alpha (HV69-70) and Delta (del156-157/R158G), normalized by N gene concentrations in wastewater, averaged over the previous 14 days. No data are shown for dates for which no measurements were made within the previous 14 days.

Supplementary Files

This is a list of supplementary files associated with this preprint. [Click to download.](#)

- [SupplementalVariantsinWastewaterSubmit.docx](#)