Chronic shedding of a SARS-CoV-2 Alpha variant in wastewater

Michael J. Conway[1,5*], Hannah Yang[1], Lauren A. Revord[1], Michael P. Novay[1], Rachel J. Lee[1], Avery S. Ward[1], Jackson D. Abel[1], Maggie R. Williams[4,5], Rebecca L. Uzarski[3] and Elizabeth W. Alm[2,5]

Abstract

Background Central Michigan University (CMU) participated in a state-wide SARS-CoV-2 wastewater monitoring program since 2021. Wastewater samples were collected from on-campus sites and nine off-campus wastewater treatment plants servicing small metropolitan and rural communities. SARS-CoV-2 genome copies were quantified using droplet digital PCR and results were reported to the health department.

Results One rural, off-campus site consistently produced higher concentrations of SARS-CoV-2 genome copies. Samples from this site were sequenced and contained predominately a derivative of Alpha variant lineage B.1.1.7, detected from fall 2021 through summer 2023. Mutational analysis of reconstructed genes revealed divergence from the Alpha variant lineage sequence over time, including numerous mutations in the Spike RBD and NTD.

Conclusions We discuss the possibility that a chronic SARS-CoV-2 infection accumulated adaptive mutations that promoted long-term infection. This study reveals that small wastewater treatment plants can enhance resolution of rare events and facilitate reconstruction of viral genomes due to the relative lack of contaminating sequences.

Keywords SARS-CoV-2, Wastewater, Surveillance, Cryptic, Chronic, Evolution

Introduction

Wastewater surveillance became an important public health tool during the COVID-19 pandemic. Wastewater surveillance programs identified outbreaks within communities and individual buildings and it is increasingly being used to detect variants of concern [1–11]. The goal of wastewater surveillance is to provide data to public health agencies so that they may make informed decisions regarding mitigation strategies such as physical distancing, masking, business closures, and distribution of resources such as prophylactic vaccines.

The State of Michigan Department of Health and Human Services (MDHHS) initiated a wastewater surveillance program in 2021. The program included partnerships between academic laboratories and regional public health departments that spanned large and small metropolitan areas and rural areas in both the lower and upper peninsulas. Central Michigan University (CMU) formed a partnership with the Central Michigan District Health Department (CMDHD). This partnership provided an opportunity to look at the dynamics of SARS-CoV-2 at a public university and in the surrounding small metropolitan and rural communities [12]. We identified ten on-campus sewer sites and nine off-campus wastewater treatment plants (WWTPs) to sample on a weekly

*Correspondence:
Michael J. Conway
michael.conway@cmich.edu
[1] Foundational Sciences, Central Michigan University, College of Medicine, Mt. Pleasant, MI, USA
[2] Department of Biology, Central Michigan University, Mt. Pleasant, MI, USA
[3] Department of Biology and Herbert H. and Grace A. Dow College of Health, Professions, Central Michigan University, Mt. Pleasant, MI, USA
[4] School of Engineering & Technology, Central Michigan University, Mt. Pleasant, MI, USA
[5] Institute for Great Lakes Research, Central Michigan University, Mt. Pleasant, MI, USA

Conway *et al. BMC Genomics*   (2024) 25:59

Page 2 of 10

basis. The population sizes serviced by these WWTPs ranged from as large as 35,397 to as small as 851.

Sampling began in July 2021, which was at least seven months after emergence of the Alpha variant (B.1.1.7). The Alpha variant first appeared in North America in late November 2020 and became the predominant SARS-CoV-2 variant by the end of March 2021. The Alpha variant diverged into multiple lineages, including B.1.1.7-derivatives like Q.3. The Q.3 lineage is present in 5,422 clinical sequences worldwide that were uploaded to the NCBI database, and a positive sample was first collected on 7-11-20 (The dates used to describe samples are formatted using the American system (MM-DD-YY)). The Q.3 lineage was detected in clinical samples from Michigan 16 times between 2-18-21 and 7-9-21.

It became clear that our smallest WWTP (estimated population served: 851) consistently produced higher concentrations of SARS-CoV-2 genome copies. Samples taken from this site from 2021 to 2023 were sequenced and many contained sequences that corresponded to an Alpha variant lineage. These sequences also accumulated novel mutations over time, including previously described mutations found only in cryptic lineages derived from wastewater and not in circulating clinical samples [13, 14]. In this manuscript, a cryptic mutation is a mutation previously identified in cryptic lineages, which were previously identified in wastewater but not clinical samples. It's important to highlight that many of the mutations previously identified in cryptic lineages have since been identified in clinical samples. We hypothesize that an individual was chronically infected with an Alpha variant lineage for 20–28 months. During this time, the virus adapted by accumulating novel mutations, which included previously described cryptic mutations [13, 14]. Importantly, we found that the earliest sample corresponded to Alpha variant lineage Q.3, which closely aligned with clinical sequences reported in summer and fall 2021; however, the sequence diverged over time and accumulated novel mutations.

These data reveal that wastewater surveillance in small metropolitan and rural communities provide an opportunity to identify novel isolates and reconstruct genes due to lower contamination with unrelated sequences. These data also suggest that humans and other animals can chronically shed SARS-CoV-2 over many months, which is associated with accumulation of adaptive mutations. Mutations associated with chronic infection may be useful to identify individuals who are chronically infected and to drive selection of appropriate therapeutics.

Materials and methods

Selection of sample sites

Central Michigan University (CMU) is a public research university in the City of Mt. Pleasant, Isabella County, Michigan, with an average population during the 2021–2022 academic year of 13,684 students and staff. Ten sample sites were selected on campus that collected wastewater downstream from most campus buildings, including residential halls, apartments, and academic/administrative buildings. The waste stream at these sites includes a mixture of wastewater from CMU and upstream residential areas in the City of Mt. Pleasant. Nine off-campus sites throughout the jurisdictions of the Central Michigan District Health Department (CMDHD) and Mid-Michigan District Health Department (MMDHD) were selected [12], which included the City of Mt. Pleasant, Union Township, City of Alma, City of Clare, City of Evart, three Houghton Lake townships, and Village of Marion wastewater treatment plants (WWTPs). These locations represent various land uses and population densities including urban, rural, and suburban areas, providing a large footprint of SARS CoV-2 virus shedding in Central Michigan.

Wastewater collection

Since July 2021, wastewater samples (500–1000 mL) were collected once each week on either Monday or Tuesday from ten sanitary sewer sites and nine WWTP influent streams (after grit removal). Sanitary sewer grab samples consisted of wastewater flowing from university dormitories and buildings and the surrounding community. Influent to WWTPs were collected as grab samples or 24-hour composite samples [12]. Samples were held at 4 °C no more than 48 h before analysis.

Virus concentration and RNA extraction

The protocol described by Flood et al. 2021 and adopted by the Michigan wastewater surveillance network was used to concentrate virus from samples and extract viral RNA [12, 15]. Briefly, 100 mL wastewater or water as a negative control was mixed with 8% (w/v) molecular biology grade PEG 8000 (Promega Corporation, Madison WI) and 0.2 M NaCl (w/v). The sample was mixed slowly on a magnetic stirrer at 4 °C for 2–16 h. Following overnight incubation, samples were centrifuged at 4,700×g for 45 min at 4 °C. The supernatant was then removed, and the pellet was resuspended in the remaining liquid, which ranged from 1 to 3 mL. All sample concentrates were aliquoted and stored at -80 °C until further processing. Viral RNA was extracted from concentrated wastewater using the Qiagen QIAmp Viral RNA Minikit according to the manufacturer's protocol with previously published modifications (Qiagen,

Germany) [15]. In this study, a total of 200 μl of concentrate was used for RNA extraction resulting in a final elution volume of 80 μl. Extracted RNA was stored at -80 °C until analysis. A wastewater negative extraction control was included. To derive recovery efficiencies for each sample site, samples were inoculated with 10^6 gene copies (GC)/mL Phi6 bacteriophage (Phi6) prior to the addition of PEG and NaCl. Wastewater samples were mixed, and a 1 mL sample was reserved and stored at -80 °C. RNA was extracted as stated above.

Detection and quantification of SARS-CoV-2

A one-step RT-ddPCR approach was used to determine the copy number/20 μL of SARS-CoV-2, and data were converted to copy number/100 mL wastewater for N1 and N2 targets using the method published by Flood et al., 2001 [15]. All the primers and probes used in this study were published previously [12]. Droplet digital PCR was performed using Bio-Rad's 1-Step RT-ddPCR Advanced kit with a QX200 ddPCR system (Bio-Rad, CA, USA). Each reaction contained a final concentration of 1 × Supermix (Bio-Rad, CA, USA), 20 U μL^{-1} reverse transcriptase (RT) (Bio-Rad, CA, USA), 15 mM DTT, 900 nmol l^{-1} of each primer, 250 nmol l^{-1} of each probe, 1 μL of molecular grade RNAse-free water, and 5.5 μL of template RNA for a final reaction volume of 22 μL [12, 15–17]. RT was omitted for DNA targets. Droplet generation was performed by microfluidic mixing of 20 μL of each reaction mixture with 70 μL of droplet generation oil in a droplet generator (Bio-Rad, CA, USA) resulting in a final volume of 40 μL of reaction mixture-oil emulsions containing up to 20,000 droplets with a minimum droplet count of > 9000. The resulting droplets were then transferred to a 96-well PCR plate that was heat-sealed with foil and placed into a C1000 96-deep-well thermocycler (Bio-Rad, CA, USA) for PCR amplification using the following parameters: 25 °C for 3 min, 50 °C for 1 h, 95 °C for 10 min, followed by 40 cycles of 95 °C for 30 s and 60 °C for 1 min with ramp rate of 2 °C/s 1 followed by a final cycle of 98 °C for 10 min. Following PCR thermocycling, each 96-well plate was transferred to a QX200 Droplet Reader (Bio-Rad, CA, USA) for the concentration determination through the detection of positive droplets containing each gene target by spectrophotometric detection of the fluorescent probe signal. All analyses were run in triplicate for each marker. To derive recovery efficiencies for each sample site, Phi6-spiked pre- and post-PEG concentration RNA samples were used to quantify Phi6 copy number using the previously published primers and probes [12]. The degree of PCR inhibition was also quantified in each sample by spiking 10 μL of 10^5 GC/ml

Phi6 in a sample's Buffer AVL, including positive controls that lacked wastewater.

Data analysis

All SARS-CoV-2 gene data were converted from GC per 20 μL reaction to GC per 100 mL wastewater sample before analysis [12, 15]. Non-detects (ND) were assigned their individual sample's limit of detection for the purposes of data reporting, although any weekly on-campus or off-campus samples whose values matched the theoretical limit of detection were removed prior to statistical analysis. The limit of detection was calculated for each individual sample based on both the molecular assays' theoretical detection limits (i.e., 3 positive droplets for RT-ddPCR; the lowest standard curve concentration for RT-qPCR) and the concentration factor of each processing method examined. All wastewater data were reported to MDHHS and uploaded to the Michigan COVID-19 Sentinel Wastewater Epidemiological Evaluation Project (SWEEP) dashboard (https://www.michigan.gov/coronavirus/stats/wastewater-surveillance/dashboard/sentinel-wastewater-epidemiology-evaluation-project-sweep).

Sequencing

RNA was shipped to GT Molecular (Fort Collins, CO) on dry ice. Library preparation was done using GT Molecular's proprietary method, which utilized ARTIC 4.1 primers for SARS-CoV-2 amplicon generation (https://artic.network/ncov-2019). Amplicons were pooled and sequenced on a Miseq using 2×150 bp reads. FASTQ files were analyzed using GT Molecular's bioinformatics pipeline, and variant-calling was performed using a modified and proprietary version of Freyja [18]. FASTQ files for each sample listed in Table 1 are available in the NCBI SRA database (Submission ID: SUB13897431; BioProject ID: PRJNA1027333).

Spike reconstruction and identification of novel mutations

FASTQ files from 11-9-21, 9-12-22, 4-24-23, and 5-1-23 contained reads that spanned the entire Spike protein, they lacked contamination with other variants of concern based on variant calling, and they had high relative abundance of the Alpha variant lineage B.1.1.7 derivative. This allowed for reconstruction of a consensus Spike gene for each of the above wastewater samples. Specifically, we uploaded FASTA-formatted .txt files into Galaxy (https://usegalaxy.org/) that represented the SARS-CoV-2 reference Spike gene. We then uploaded each of the paired-end FASTQ files for each wastewater sample. The Bowtie2 program was used to map reads against each reference sequence, creating individual .bam files per sample. The default setting was used for analysis. The Convert Bam program was then used to

Conway *et al. BMC Genomics* (2024) 25:59

Page 4 of 10

Table 1 GT molecular variant calling

Location code	Sample date	VOC (%)[a]	Lineage(s) (%)[b]
VM	9-21-21	Delta (98.4)	AY.25.1 (84.8)
VM	10-26-21	Alpha (94.2)	Q.3 (94.2)
VM	11-9-21	Alpha (94.9)	Q.4 (94.7)
VM	3-14-22	Omicron (93.8)	XBB.2.3.11 (74.3)
VM	4-25-22	Omicron (95.1)	XBB.1.5.81 (17.7)
			XBB.1.16.1 (17.5)
			XBB.1.5.10 (15.0)
			EG.5 (7.7)
			XBB.1.5.72 (7.4)
			GJ.1.2 (5.9)
			FD.1.1 (5.1)
VM	9-12-22	Alpha (97.3)	Q.4 (96.9)
VM	3-13-23	Alpha (98.0)	Q.4 (65.5)
VM	3-27-23	Omicron (27.1)	Q.4 (47.1)
		Alpha (47.1)	B.1.1.28 (16.2)
			XBB.1.5.49 (7.9)
VM	4-24-23	Alpha (97.9)	Q.4 (89.7)
VM	5-1-23	Alpha (97.4)	Q.4 (97.1)
VM	5-15-23	Alpha (96.6)	Q.4 (93.8)
C6	1-2-22	Omicron (89.9)	BA.1.1 (82.5)
		Delta (7.6)	B.1.617.2 (7.3)
CA	6-6-22	Omicron (99.8)	BG.5 (99.8)
MP	2-20-23	Omicron (98.4)	XBB.1.5 (45.1)
CL	2-28-23	Omicron (99.1)	DT.2 (49.9)
			BQ.1.1 (16.1)
			BQ.1.1.37 (16.1)
			BQ.1.1.52 (16.1)
UT	4-3-23	Omicron (86.5)	XBB.1.5 (19.6)
		Delta (7.4)	XBB.1.5.23 (15.9)
			XBB.1.28 (15.6)
			XBB.1.5.1 (8.2)
			XBB.1.5.17 (7.5)
CE	4-10-23	Alpha (89.0)	Q.4 (88.8)

[a] Relative abundance of variants of concern (VOC) as a percentage

[b] Relative abundance of VOC lineages as a percentage

convert .bam files to FASTA multiple sequence alignments. Multiple sequence alignment files were uploaded to MEGA (https://www.megasoftware.net/) and converted to amino acid sequence The consensus amino acid sequence from each of these samples was manually reconstructed and then aligned with the SARS-CoV-2 Spike reference sequence and a consensus Alpha variant lineage Q.3 sequence derived from 16 clinical samples collected in Michigan from 2-18-21 to 7-9-21. The Q.3 lineage was chosen because the earliest wastewater sample that tested positive for SARS-CoV-2 Alpha variant (i.e., 10-26-21) was Alpha variant lineage Q.3 based on the GT Molecular variant calling pipeline. Mutations that were present in wastewater samples but not

in the SARS-CoV-2 Spike reference sequence or clinical sample were characterized as novel mutations. FastQC was used to quantify the total number of reads in each FASTQ file, the total number of reads that aligned to the reference Spike, the read length, and the number of poor-quality sequences (Supplementary Table 1).

Novel and cryptic mutation hotspot analyses

We identified novel mutations as described above. Previous literature also identified cryptic sequence hotspots in SARS-CoV-2 Spike [13, 14]. We tracked the percent prevalence of novel and cryptic mutations in wastewater samples that were positive for the Alpha variant lineage. Specifically, we uploaded FASTA-formatted .txt files into Galaxy (https://usegalaxy.org/) that represented the SARS-CoV-2 reference Spike. We then uploaded each of the paired-end FASTQ files for each wastewater sample. The Bowtie2 program was used to map reads against the reference sequence. The default setting was used for analysis. The Convert Bam program was then used to convert .bam files to FASTA multiple sequence alignments. Multiple sequence alignment files were uploaded to MEGA (https://www.megasoftware.net/) and converted to amino acid sequence for open-reading frame analysis. Novel and cryptic mutations were identified manually, and the column of reads were copied and pasted into Excel. The column was selected, and the Analyze Data tool was selected to calculate the percent prevalence of the novel and cryptic mutations. This was repeated for each novel and cryptic mutations across all samples positive for Alpha variant lineage and the percent prevalence data was represented as heatmaps. Novel mutations present in the 2021, 2022, and 2023 consensus Spike proteins were mapped onto the furin cleaved spike protein of SARS-CoV-2 with one RBD erect using UCSF Chimera [19]. This atomic structure was selected because it had the greatest resolution of each amino acids across the Spike protein and allowed mapping of most novel mutations.

Results

Chronic shedding of an Alpha variant lineage at a rural WWTP

Wastewater samples were collected between July 2021 and June 2023 from ten on-campus sanitary sewer sites and nine WWTP influent streams. SARS-CoV-2 genome copies per 100 mL wastewater were determined each week and reported to MDHHS. One site was notable for higher peaks of virus shedding, which culminated in a peak that was 4 logs higher than the mean for all sites, although high peaks of activity were observed since 9-21-21 (Fig. 1). In order to identify the SARS-CoV-2 variant responsible for
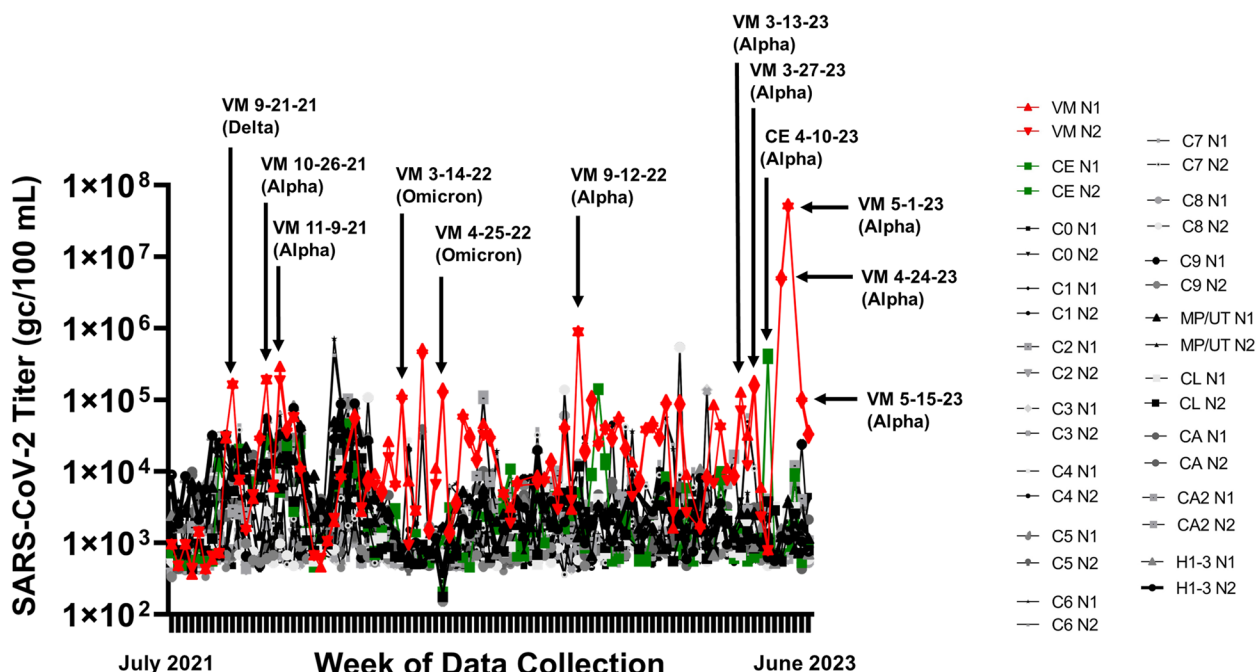
Fig. 1 SARS-CoV-2 genome copies (GC)/100 mL wastewater detected at each weekly sample site from July 2021 to June 2023. Two letter site codes and dates are shown that correspond to sequenced samples and the variant that was identified in the highest relative abundance is indicated in parentheses. The colors and shapes associated with each sample are located in the graphical legend

this activity, RNA extracted from stored wastewater concentrates was shipped to GT Molecular and their NGS and variant calling pipeline was used. RNA from the site of interest and neighboring sites were analyzed as a control. The site of interest contained high relative abundance of

Delta variant lineage AY.25.1 at the first time point tested (i.e., 9-21-21) (Fig. 1; Table 1). This corresponded to the beginning of the Delta variant wave in Central Michigan [12]. The site of interest began shedding the Alpha variant lineage during the next two time points tested (i.e.,

Table 2 SARS-CoV-2 Alpha variant Q.3/Q.4 clinical sequence from Michigan

| Accession	Organism	Pangolin	Geo location	Host	Isolation source	Collection date
OL892482	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	swab	2/18/2021
MZ158978	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	unknown	3/29/2021
MW991267	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	unknown	3/31/2021
MZ025257	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	unknown	4/4/2021
OL812274	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	swab	4/8/2021
MZ071109	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	oronasopharynx	4/12/2021
MZ131434	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	oronasopharynx	4/16/2021
OL803302	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	swab	4/23/2021
MZ416155	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	oronasopharynx	5/12/2021
MZ353922	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	oronasopharynx	5/20/2021
OK296641	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	unknown	5/24/2021
OK233076	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	oronasopharynx	5/25/2021
OK292910	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	unknown	6/12/2021
OK187847	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	unknown	6/18/2021
MZ780743	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	unknown	6/18/2021
MZ728995	SARS-CoV-2	Q.3	USA: Michigan	Homo sapiens	unknown	7/9/2021
MZ025226	SARS-CoV-2	Q.4	USA: Michigan	Homo sapiens	unknown	4/3/2021

10-26-21 and 11-9-21) (Fig. 1; Table 1). This was preceded by sequencing data from clinical samples, which revealed 16 Alpha variant lineage Q.3 isolates collected from 2-18-21 to 7-9-21 (Table 2). The site of interest had high relative abundance of Omicron variant lineages during the next two time points tested (i.e., 3-14-22 and 4-25-22) (Fig. 1; Table 1). This corresponded to the end of the first Omicron wave in Central Michigan [12]. The Alpha variant lineage became the dominant isolate in all remaining wastewater samples from the site of interest in all 2022 and 2023 samples tested, with relative abundance ranging from 47.1 to 98.0%. The Alpha variant lineage was also detected in the closest neighboring WWTP on 4-10-23, which corresponded to a large peak in virus shedding at that site (Fig. 1; Table 1). Other sites contained Omicron variant lineages BG.5, XBB.1.5, XBB.1.5.23, XBB.1.28, XBB.1.5.1, XBB.1.5.17, XBB.1.5.49, and Delta variant lineage DT.2 at varying relative abundance (Table 1).

Accumulation of novel mutations in the RBD and NTD

We reasoned that chronic shedding of SARS-CoV-2 would lead to accumulation of novel or cryptic mutations that do not align with sequences identified in most clinical and wastewater samples. Alignment of reconstructed consensus genes with the SARS-CoV-2 Spike reference gene and a consensus Alpha variant lineage clinical sequence revealed that the Spike proteins harbored 9 novel mutations in 2021, 25 novel mutations in 2022, and 38 novel mutations in 2023 (Supplemental Fig. 1). We expanded this analysis by quantifying the percent prevalence of each of the 38 novel mutations identified in the 2023 samples across all wastewater samples that were positive for the Alpha variant lineage. A heatmap showed that these mutations accumulated within the population over time, while also retaining diversity at each position (Fig. 2). This analysis was also performed using previously published cryptic mutations [13, 14].

The majority of the novel mutations accumulated in the receptor binding domain (RBD) and N-terminal domain (NTD) of the surface (a.k.a., spike) glycoprotein (Fig. 3A–F). The following mutations were identified in the RBD: A372T, R403K, K444R, V445A, G446D, Y449N, L452Q, Y453F, N460K, S477N, E484V, Q493K, Q498L, G504D, and Y505H. The following mutations were identified in the NTD: T19K, H49Y, W64R, H66Q, I68Del, T76I, V143Del, Y144Del, K147T, S151I, G181E, N196S, Y248S, and G257D. Additional mutations were identified toward the C-terminus: D571G, I587V, V772A, L828F, T941S, V1176F, K1191N, and Q1201K. Many of these mutations were not mapped onto the Spike protein due to a lack of resolution toward the C-terminus (Supplementary Fig. 1).

The closest neighboring WWTP also contained the Alpha variant lineage (Table 1). Alignment of the reconstructed
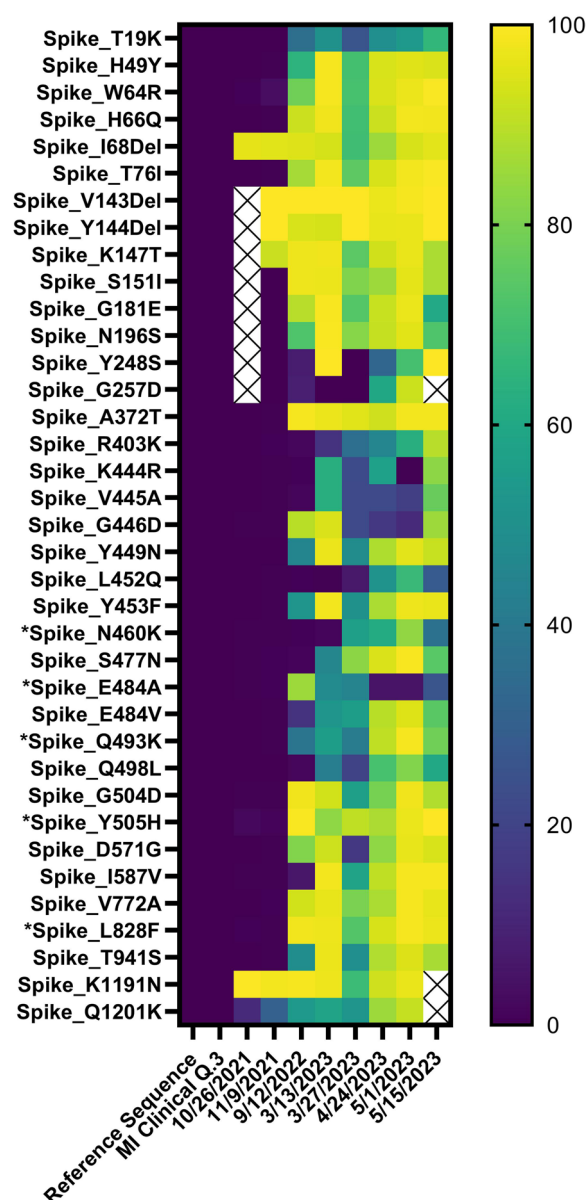
Fig. 2 Heatmap showing the percent prevalence of novel and previously identified cryptic mutations (*) in each wastewater sample that was positive for the Alpha variant lineage [13, 14]. Empty cells represent mutations that had fewer than 3 reads

Spike gene from CE 4-10-23 revealed shared mutations with reconstructed Spike genes from 11-9-21, 9-12-22, and 5-1-23, and four unique mutations: L24S, H245Y, V445A, and Y1155F (Supplementary Fig. 2). The mutations shared with the reconstructed Spike proteins from 11-9-21, 9-22-22, and 5-1-23 suggested that CE 4-10-23 shared a common ancestor with more recent isolates.
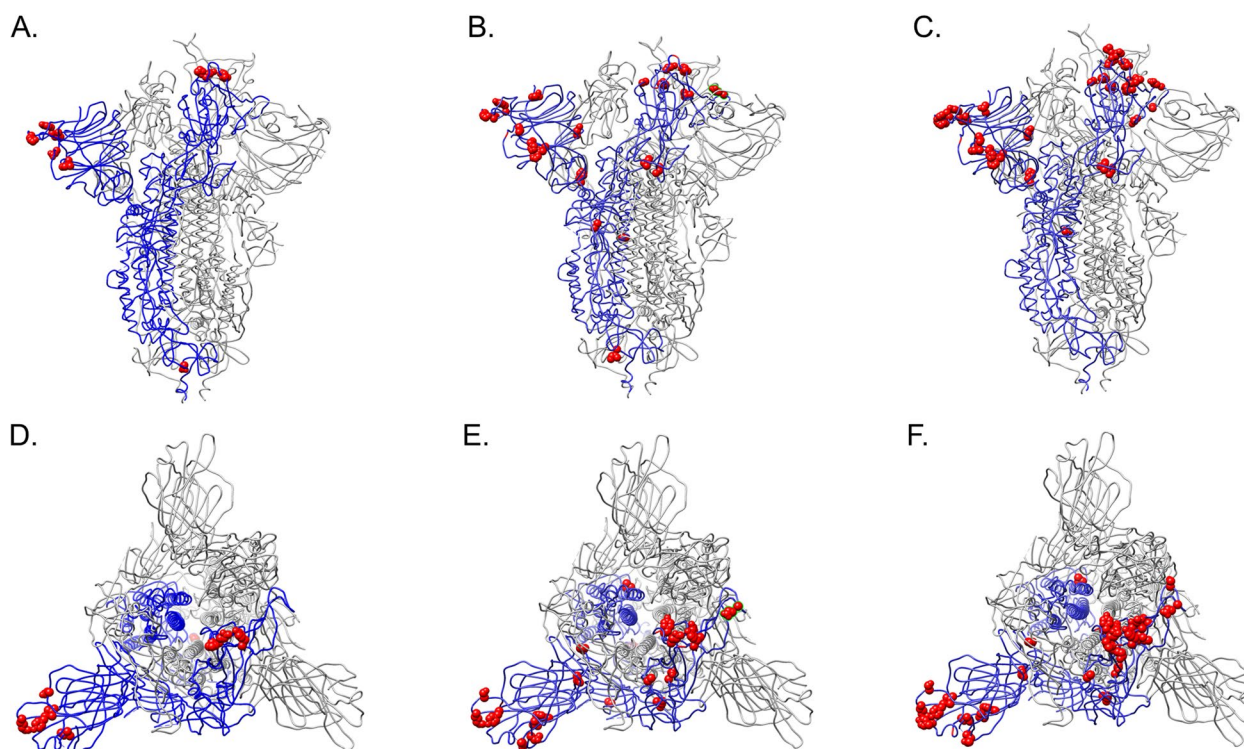
Fig. 3 Novel mutations present in the (**A**, **D**) 2021, (**B**, **E**) 2022, and (**C**, **F**) 2023 consensus Spike proteins were mapped onto the furin cleaved spike protein of SARS-CoV-2 with one RBD erect represented as red Corey-Pauling_Koltun (CPK) spheres using UCSF Chimera [19]

Discussion

Central Michigan University participated in a state-wide SARS-CoV-2 wastewater surveillance network starting July 2021. Two major waves of SARS-CoV-2 passed through the central Michigan region during the 2021/2022 academic year, which were characterized by the emergence of Delta and Omicron variants [12]. Delta and Omicron variants were preceded by the Alpha variant B.1.1.7. The SARS-CoV-2 Alpha variant B.1.1.7 wave passed through central Michigan during winter/spring 2021. Distinct Alpha variant subpopulations were also present at that time including Q.3 and Q.4 lineages.

Retrospective analysis of wastewater data revealed that one rural site produced consistently higher concentrations of SARS-CoV-2 copy numbers since September 2021. NGS sequencing revealed that this site began shedding an Alpha variant lineage by October 2021 and that this continued to at least May 2023. Clinical sequence data revealed that Alpha variant lineage Q.3/Q.4 was present in Michigan between February to July 2021. This preceded the start of wastewater surveillance in central Michigan and our first detection of the Alpha variant lineage in wastewater by 3–8 months. It is unclear how many individuals were originally infected with this lineage at the site of interest, and it is unclear how many

individuals continued to shed the virus into the sewer shed. However, due to the small population served at this rural WWTP, our November 2021 Alpha variant lineage Spike gene reconstruction likely represents a chronic infection that lasted for 3–8 months. At this stage of the chronic infection, the Alpha variant lineage already accumulated 9 novel mutations in the Spike gene.

Most of the mutations reside in the Spike RBD and NTD. These domains are critical for host receptor binding and contain key epitopes leveraged by the adaptive immune system to control and prevent repeat infection. A striking mutation that developed in 2023 was R403K. This converted the RGD receptor binding motif to KGD, which is present in SARS-CoV-1 – a historically more lethal yet less transmissible virus [20–22]. This is particularly interesting since R403 is highly conserved in SARS-CoV-2 Spike and only 294 of ~ 3.4 million sequences recorded on GSAID contained a conservative change of R403K [23]. Many other mutations have also been previously characterized. For instance, engineering the A372T mutation into SARS-CoV-2 reduced binding to ACE2 and enhanced replication in human lung cells [24]. K444R, V445A, G446D, Y449N, L452Q, N460K, S477N, and E484V (and cryptic mutation E484A) have been associated with resistance to antibody-mediated

Conway *et al. BMC Genomics*   (2024) 25:59

Page 8 of 10

neutralization, and N460K was previously observed during a persistent infection in an immunocompromised patient [14, 25–33]. E848V also reduced ACE2 binding [33]. L452Q had higher binding to soluble ACE2 [31]. Y453F increased binding to mink ACE2 [34]. Q493K increased binding to mouse ACE2 and developed in an immunocompromised patient undergoing convalescent plasma treatment [35–37]. Q498L was predicted to lower stability of Spike and ACE2 interaction but no studies are available to confirm this prediction [38]. G504D is associated with immune evasion; however, the G504D substitution is rarely observed in SARS-CoV-2 strains, with a mutant rate below 0.002% [29, 39]. Y505H was in all lineages of the Omicron variant suggesting that it enhanced immune evasion and receptor binding [40].

In the NTD, H49Y impacts Spike structure and influences binding of several antiviral compounds and increased resistance to vaccine sera [25, 41]. T76I increases infectivity in the Lambda variant, although it is suggested that it behaves as a compensatory mutation [31]. T76I also effects antibody binding and immune escape [42]. V143Del is present in Omicron variants suggesting that it enhanced immune evasion and receptor binding [43]. Y144Del may play a role in ACE2 receptor binding or neutralizing antibody escape and deletions in this region were identified in immunocompromised patients [37, 44, 45]. T19K, W64R, H66Q, I68Del, K147T, S151I, G181E, N196S, Y248S, and G257D substitutions have not been characterized. Toward the C-terminus, L828F is a highly prevalent cryptic mutation, which has an unknown origin, although likely due to shedding from chronically infected humans or animals [13, 14]. D571G, I587V, V772A, T941S, V1176F, K1191N, and Q1201K substitutions have not been characterized.

Collectively, the mutations that have accumulated in the Spike gene are likely a response to the host's innate and adaptive immune systems, and perhaps due to long-term persistence in an immunosuppressed patient and adaptation to any prophylactic or targeted drugs used to clear the infection [37, 45–47]. The presence of previously identified cryptic mutations suggest that these mutations may predict a chronic infection. The goal of this work is not to identify the person(s) responsible for chronically shedding this virus. However, we would like to highlight the potential of wastewater surveillance and possibly fecal testing for identification of long COVID. This would be particularly useful if convergent mutations emerge during a chronic infection that are predictive of this condition. The spectrum of mutations might guide appropriate selection of antivirals and antibody-based therapies. Additional mutations outside of the Spike gene were also present in these samples but not analyzed for this manuscript. It is likely that mutations outside of the Spike gene are also important to facilitate chronic infection.

In summary, these data support that an individual can be chronically infected with SARS-CoV-2 over many months and possibly a few years. During this time, SARS-CoV-2 can accumulate many mutations in the Spike gene, which concentrate in the RBD and NTD. Further research is needed to determine if these mutations are predictive of chronic infection and if they can be used as a biomarker in individuals with Long COVID and leveraged to tailor selection or development of pharmaceutical therapies. Additionally, this study shows that small WWTPs can enhance the resolution of rare biological events and allow for total reconstruction of viral genes and their corresponding proteins.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-09977-7.

Additional file 1: Supplemental Table 1. MEGA and FastQC analyses of wastewater samples positive for an Alpha variant lineage.

Additional file 2: Supplemental Figure 1. Clustal Omega alignment of reconstructed 2021, 2022, and 2023 Spike proteins with reference and consensus Alpha variant Q.3 clinical sequence Spike proteins.

Additional file 3: Supplemental Figure 2. Clustal Omega alignment of reconstructed 2021, 2022, and 2023 Spike proteins with reconstructed Spike protein from CE 4-10-23. Mutations unique to CE-4-10-23 are highlighted in grey.

Authors' contributions

MJC wrote the manuscript and directed the wastewater monitoring activities, HY and LAR analyzed FASTQ data and developed the heatmap, ASW and JDA performed all wastewater monitoring activities including submission of samples for NGS, MRW supported data analysis and submission to the health department and manuscript revision, RLU served as a liaison between wastewater treatment plants and student research assistants, and EWA assisted in data analysis and manuscript revision. MPN and RJL reanalyzed data and revised the manuscript during the peer review process.

Availability of data and materials

FASTQ files for each sample listed in Table 1 are available in the NCBI SRA database (Submission ID: SUB13897431; BioProject ID: PRJNA1027333).

Conway *et al. BMC Genomics*   (2024) 25:59

Page 9 of 10

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References

1. Anderson-Coughlin BL, Shearer AEH, Omar AN, Litt PK, Bernberg E, Murphy M, et al. Coordination of SARS-CoV-2 wastewater and clinical testing of university students demonstrates the importance of sampling duration and collection time. Sci Total Environ. 2022;830:154619.
2. Betancourt WQ, Schmitz BW, Innes GK, Prasek SM, Pogreba Brown KM, Stark ER, et al. COVID-19 containment on a college campus via wastewater-based epidemiology, targeted clinical testing and an intervention. Sci Total Environ. 2021;779:146408.
3. Corchis-Scott R, Geng Q, Seth R, Ray R, Beg M, Biswas N, et al. Averting an outbreak of SARS-CoV-2 in a University Residence hall through wastewater surveillance. Microbiol Spectr. 2021;9(2):e0079221.
4. Gibas C, Lambirth K, Mittal N, Juel MAI, Barua VB, Roppolo Brazell L, et al. Implementing building-level SARS-CoV-2 wastewater surveillance on a university campus. Sci Total Environ. 2021;782:146749.
5. Jarvie M, Reed M, Southwell M, Wright B, Nguyen D, Ngoc T. Thi. RT-ddPCR Wastewater monitoring of COVID-19 across the Eastern upper Peninsula of Michigan. SSRN. 2022.
6. Scott LC, Aubee A, Babahaji L, Vigil K, Tims S, Aw TG. Targeted wastewater surveillance of SARS-CoV-2 on a university campus for COVID-19 outbreak detection and mitigation. Environ Res. 2021;200:111374.
7. Wong TE, Thurston GM, Barlow N, Cahill ND, Carichino L, Maki K, et al. Evaluating the sensitivity of SARS-CoV-2 infection rates on college campuses to wastewater surveillance. Infect Dis Model. 2021;6:1144–58.
8. Rondeau NC, Rose OJ, Alt ER, Ariyan LA, Elikan AB, Everard JL, et al. Building-level detection threshold of SARS-CoV-2 in wastewater. Microbiol Spectr. 2023;11:e0292922.
9. Chua FJD, Kim SY, Hill E, Cai JW, Lee WL, Gu X, et al. Co-incidence of BA.1 and BA.2 at the start of Singapore's Omicron wave revealed by community and university campus wastewater surveillance. Sci Total Environ. 2023;875:162611.
10. Vo V, Tillett RL, Papp K, Chang CL, Harrington A, Moshi M, et al. Detection of the Omicron BA.1 variant of SARS-CoV-2 in Wastewater from a Las Vegas Tourist area. JAMA Netw Open. 2023;6(2):e230550.
11. Pico-Tomàs A, Mejías-Molina C, Zammit I, Rusiñol M, Bofill-Mas S, Borrego CM, et al. Surveillance of SARS-CoV-2 in sewage from buildings housing residents with different vulnerability levels. Sci Total Environ. 2023;872:162116.
12. Conway MJ, Kado S, Kooienga BK, Sarette JS, Kirby MH, Marten AD, et al. SARS-CoV-2 wastewater monitoring in rural and small metropolitan communities in Central Michigan. Sci Total Environ. 2023;894:165013.
13. Smyth DS, Trujillo M, Gregory DA, Cheung K, Gao A, Graham M, et al. Tracking cryptic SARS-CoV-2 lineages detected in NYC wastewater. Nat Commun. 2022;13(1):635.
14. Gregory DA, Trujillo M, Rushford C, Flury A, Kannoly S, San KM, et al. Genetic diversity and evolutionary convergence of cryptic SARS-CoV-2 lineages detected via wastewater sequencing. PLoS Pathog. 2022;18(10):e1010636.
15. Flood MT, D'Souza N, Rose JB, Aw TG. Methods evaluation for rapid concentration and quantification of SARS-CoV-2 in raw wastewater using droplet digital and quantitative RT-PCR. Food Environ Virol. 2021;13(3):303–15.
16. Lu X, Wang L, Sakthivel SK, Whitaker B, Murray J, Kamili S, et al. US CDC real-time reverse transcription PCR panel for detection of severe acute respiratory syndrome coronavirus 2. Emerg Infect Dis. 2020;26(8):1654–65.
17. Lee HW, Lee HM, Yoon SR, Kim SH, Ha JH. Pretreatment with propidium monoazide/sodium lauroyl sarcosinate improves discrimination of infectious waterborne virus by RT-qPCR combined with magnetic separation. Environ Pollut. 2018;233:306–14.
18. Karthikeyan S, Levy JI, De Hoff P, Humphrey G, Birmingham A, Jepsen K, et al. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. Nature. 2022;609(7925):101–8.
19. Wrobel AG, Benton DJ, Xu P, Roustan C, Martin SR, Rosenthal PB, et al. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. Nat Struct Mol Biol. 2020;27(8):763–7.
20. Luan J, Lu Y, Gao S, Zhang L. A potential inhibitory role for integrin in the receptor targeting of SARS-CoV-2. J Infect. 2020;81(2):318–56.
21. Makowski L, Olson-Sidford W, Weisel WJ. Biological and Clinical consequences of integrin binding via a Rogue RGD Motif in the SARS CoV-2 spike protein. Viruses. 2021;13(2):146.
22. Norris EG, Pan XS, Hocking DC. Receptor-binding domain of SARS-CoV-2 is a functional αv-integrin agonist. J Biol Chem. 2023;299(3):102922.
23. Zech F, Schniertshauer D, Jung C, Herrmann A, Cordsmeier A, Xie Q, et al. Spike residue 403 affects binding of coronavirus spikes to human ACE2. Nat Commun. 2021;12(1):6855.
24. Kang L, He G, Sharp AK, Wang X, Brown AM, Michalak P, et al. A selective sweep in the spike gene has driven SARS-CoV-2 human adaptation. Cell. 2021;184(17):4392-400.e4.
25. Wang Q, Ye SB, Zhou ZJ, Song AL, Zhu X, Peng JM, et al. Key mutations in the spike protein of SARS-CoV-2 affecting neutralization resistance and viral internalization. J Med Virol. 2023;95(1):e28407.
26. Ku Z, Xie X, Davidson E, Ye X, Su H, Menachery VD, et al. Molecular determinants and mechanism for antibody cocktail preventing SARS-CoV-2 escape. Nat Commun. 2021;12(1):469.
27. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe. 2021;29(3):463-476.e6.
28. Liu Z, VanBlargan LA, Bloyet LM, Rothlauf PW, Chen RE, Stumpf S, et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. Cell Host Microbe. 2021;29(3):477-88.e4.
29. VanBlargan LA, Adams LJ, Liu Z, Chen RE, Gilchuk P, Raju S, et al. A potently neutralizing SARS-CoV-2 antibody inhibits variants of concern by utilizing unique binding residues in a highly conserved epitope. Immunity. 2021;54(10):2399-416.e6.
30. Tortorici MA, Beltramello M, Lempp FA, Pinto D, Dang HV, Rosen LE, et al. Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. Science. 2020;370(6519):950–7.
31. Kimura I, Kosugi Y, Wu J, Zahradnik J, Yamasoba D, Butlertanaka EP, et al. The SARS-CoV-2 Lambda variant exhibits enhanced infectivity and immune resistance. Cell Rep. 2022;38(2):110218.
32. Wang Q, Iketani S, Li Z, Guo Y, Yeh AY, Liu M, et al. Antigenic characterization of the SARS-CoV-2 Omicron subvariant BA.2.75. Cell Host Microbe. 2022;30(11):1512-7.e4.
33. Greaney AJ, Starr TN, Barnes CO, Weisblum Y, Schmidt F, Caskey M, et al. Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. Nat Commun. 2021;12(1):4196.
34. Ren W, Lan J, Ju X, Gong M, Long Q, Zhu Z, et al. Mutation Y453F in the spike protein of SARS-CoV-2 enhances interaction with the mink ACE2 receptor for host adaption. PLoS Pathog. 2021;17(11):e1010053.
35. Huang K, Zhang Y, Hui X, Zhao Y, Gong W, Wang T, et al. Q493K and Q498H substitutions in Spike promote adaptation of SARS-CoV-2 in mice. EBioMedicine. 2021;67:103381.
36. Zhang X, Wu S, Wu B, Yang Q, Chen A, Li Y, et al. SARS-CoV-2 Omicron strain exhibits potent capabilities for immune evasion and viral entrance. Signal Transduct Target Ther. 2021;6(1):430.
37. Chen L, Zody MC, Di Germanio C, Martinelli R, Mediavilla JR, Cunningham MH, et al. Emergence of multiple SARS-CoV-2 antibody escape variants in an immunocompromised host undergoing convalescent plasma treatment. mSphere. 2021;6(4):e0048021.

38. Mishra PM, Anjum F, Uversky VN, Nandi CK. SARS-CoV-2 spike mutations modify the interaction between virus Spike and human ACE2 receptors. Biochem Biophys Res Commun. 2022;620:8–14.

39. Chen Y, Zhao X, Zhou H, Zhu H, Jiang S, Wang P. Broadly neutralizing antibodies to SARS-CoV-2 and other human coronaviruses. Nat Rev Immunol. 2023;23(3):189–99.

40. Zhou Y, Zhi H, Teng Y. The outbreak of SARS-CoV-2 omicron lineages, immune escape, and vaccine effectivity. J Med Virol. 2023;95(1):e28138.

41. Sixto-López Y, Correa-Basurto J, Bello M, Landeros-Rivera B, Garzón-Tiznado JA, Montaño S. Structural insights into SARS-CoV-2 spike protein and its natural mutants found in Mexican population. Sci Rep. 2021;11(1):4659.

42. Ou J, Lan W, Wu X, Zhao T, Duan B, Yang P, et al. Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events. Signal Transduct Target Ther. 2022;7(1):138.

43. Cao Y, Wang J, Jian F, Xiao T, Song W, Yisimayi A, et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. Nature. 2022;602(7898):657–63.

44. Shen L, Bard JD, Triche TJ, Judkins AR, Biegel JA, Gai X. Rapidly emerging SARS-CoV-2 B.1.1.7 sub-lineage in the United States of America with spike protein D178H and membrane protein V70L mutations. Emerg Microbes Infect. 2021;10(1):1293–9.

45. Goes LR, Siqueira JD, Garrido MM, Alves BM, Cicala C, Arthos J, et al. Evidence of recurrent selection of mutations commonly found in SARS-CoV-2 variants of concern in viruses infecting immunocompromised patients. Front Microbiol. 2022;13:946549.

46. Kemp SA, Collier DA, Datir RP, Ferreira I, Gayed S, Jahun A, et al. SARS-CoV-2 evolution during treatment of chronic infection. Nature. 2021;592(7853):277–82.

47. Ahmadi AS, Zadheidar S, Sadeghi K, Nejati A, Salimi V, Hajiabdolbaghi M, et al. SARS-CoV-2 intrahost evolution in immunocompromised patients in comparison with immunocompetent populations after treatment. J Med Virol. 2023;95(6):e28877.

Publisher's Note