



OPEN Wastewater sequencing from a rural community enables identification of widespread adaptive mutations in a SARS-CoV-2 alpha variant

Michael J. Conway^{1,5}✉, Michael P. Novay¹, Carson M. Pusch¹, Avery S. Ward¹, Jackson D. Abel¹, Maggie R. Williams^{4,5}, Rebecca L. Uzarski³ & Elizabeth W. Alm^{2,5}

Central Michigan University (CMU) participated in a state-wide wastewater monitoring program starting in 2021. One rural site consistently produced higher concentrations of SARS-CoV-2 genome copies. Samples from this site were sequenced retrospectively and exclusively contained a derivative of Alpha variant lineage B.1.1.7 that shed from the same site for 20–28 months. Complete reconstruction of each SARS-CoV-2 open reading frame (ORF) and alignment to an early B.1.1.7 clinical isolate identified novel mutations that were selected in non-structural (nsp1, nsp2, nsp3, nsp4, nsp5/3CLpro, nsp6, RdRp, nsp15, nsp16, ORF3a, ORF6, ORF7a, and ORF7b) and structural genes (Spike, M, and N). These were rare mutations that have not accumulated in clinical samples worldwide. Mutational analysis revealed divergence from the reference Alpha variant lineage sequence over time. We present each of the mutations on available structural models and discuss the potential role of these mutations during a chronic infection. This study further supports 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 and identifies mutations that may be associated with chronic infections.

Wastewater monitoring has become a firmly established public health tool since the COVID-19 pandemic. Wastewater monitoring programs have helped identify potential outbreaks within communities and individual buildings, they can track variants of concern, and they are being leveraged for new emerging infectious diseases^{1–12}. The goal of wastewater monitoring is to provide complementary data to public health agencies so that they can make informed decisions to mitigate infectious disease transmission.

The State of Michigan Department of Health and Human Services (MDHHS) initiated a wastewater monitoring 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 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 regional public university and in the surrounding small metropolitan and rural communities¹³. We identified ten on-campus sewer sites and nine off-campus wastewater treatment plants (WWTPs) to sample on a weekly basis.

Sampling began in July 2021, which was at least seven months after emergence of the Alpha variant (B.1.1.7) in Michigan. 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 2021¹⁴. 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 retrospectively sequenced and many contained sequences that corresponded to an Alpha variant lineage B.1.1.7. We reconstructed the Spike gene and identified 37 mutations that accumulated in the RBD and NTD¹⁵. Here, we use the same set of data to provide a complete

¹Foundational Sciences, Central Michigan University College of Medicine, Mount Pleasant, MI 48859, USA.

²Department of Biology, Central Michigan University, Mount Pleasant, MI, USA. ³Department of Biology and Herbert H. and Grace A. Dow College of Health Professions, Central Michigan University, Mount Pleasant, MI, USA.

⁴School of Engineering & Technology, Central Michigan University, Mount Pleasant, MI, USA. ⁵Institute for Great Lakes Research, Central Michigan University, Mount Pleasant, MI, USA. ✉email: michael.conway@cmich.edu

reconstruction of each SARS-CoV-2 open reading frame (ORF). Alignment of each ORF to an early B.1.1.7 clinical isolate identified novel mutations that were selected in non-structural (nsp1, nsp2, nsp3, nsp4, nsp5/3 CLpro, nsp6, RdRp, nsp15, nsp16, ORF3a, ORF6, ORF7a, and ORF7b) and structural genes (Spike, M, and N). Each of these mutations were present in less than 2% of clinical samples present in GenBank and the sequence read archive (SRA) from Dec 2023 to Jun 2024 and all but three mutations were present in less than 2% of all sequences in the GISAID worldwide database. These were rare mutations, yet three were previously associated with immunodeficiency, adaptation to remdesivir, and reinfection of a hospital worker^{16–20}. Temporal mutational analysis revealed divergence from the reference Alpha variant lineage sequence over time and these changes were confirmed in sequencing reads that contained lineage-defining mutations. Each mutation was mapped onto available structural models, and we discuss the potential significance of these changes during a chronic SARS-CoV-2 infection.

This manuscript provides further support that wastewater monitoring in small metropolitan and rural communities is an opportunity to identify novel variants and reconstruct whole genomes due to lower contamination with unrelated sequences. It is important to note that the reconstruction strategy that was used will incorporate contaminating variant sequences if they are present. These data also support that humans can chronically shed SARS-CoV-2 for close to two years. Considering the low prevalence of these mutations in clinical samples, chronic shedding of SARS-CoV-2 is likely a rare event that leads to accumulation of adaptive mutations. Identifying mutations associated with chronic infection may be useful to diagnose individuals who have persistent disease and to assist in the selection of appropriate treatment.

Materials and methods

Selection of sample sites

The materials and methods are largely identical to those described in our previous manuscript but used to analyze each open reading frame across the SARS-CoV-2 genome¹⁵. 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 selected¹³, 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 Michigan¹⁵.

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. Samples were held at 4 °C no more than 48 h before analysis^{13,15}.

Virus concentration and RNA extraction

The protocol described by Flood et al. 2021 and adopted by the Michigan wastewater monitoring network was used to concentrate virus from samples and extract viral RNA^{13,15,21}. 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)²¹. 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 10⁶ 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 above¹⁵.

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., 2001²¹. All the primers and probes used in this study were published previously^{13,15}. 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^{13,21–23}. 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 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^{13,15}. The degree of PCR inhibition was also quantified in each sample by spiking 10 µL of 10⁵ GC/ml Phi6 in a sample's Buffer AVL, including positive controls that lacked wastewater¹⁵.

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^{13,15,21}. 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>)¹⁵. Weekly case data per county was acquired from the MDHHS Michigan COVID-19 Cases and Deaths dashboard (<https://www.michigan.gov/coronavirus>).

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 Freyja²⁴. FASTQ files for each sample listed are available in the NCBI SRA database (Submission ID: SRP465974; BioProject ID: PRJNA1027333)¹⁵. Sequencing reads that contained Alpha variant-defining mutations were also curated for downstream analysis. Specifically, sequencing reads from each sample were aligned to the SARS-CoV-2 reference genome (Hu-1) and lineage-defining single-nucleotide variants (SNVs), referred to as UShER barcodes, were used to identify reads supporting the B.1.1.7 and Q.3 Alpha variants. These barcodes, as implemented in the Freyja pipeline represent unique sets of mutations associated with specific SARS-CoV-2 lineages and are derived from the UShER phylogenetic framework. Reads containing one or more of the relevant barcode SNVs were selectively retained, while those lacking all lineage-defining markers were excluded. This approach generated a curated dataset enriched for B.1.1.7 and Q.3 Alpha variant-specific reads, while preserving original mapping coordinates and coverage profiles.

Complete reconstruction and identification of novel mutations

FASTQ files from 10–26-21 (SAMN37791375), 11-9-21 (SAMN37791376), 9–12-22 (SAMN37791379), 3–13-23 (SAMN37791380), 4–24-23 (SAMN37791382), and 5-1-23 (SAMN37791383) contained reads that spanned each SARS-CoV-2 open reading frame (ORF), 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 (Table 1)¹⁵. This allowed for reconstruction of consensus genes for each of the above wastewater samples. Specifically, we uploaded FASTA-formatted.txt files into Galaxy (<https://usegalaxy.org/>) that represented each reference gene. Reference genes were constructed from an early consensus Alpha variant lineage B.1.1.7 Michigan clinical isolate submitted on 1–26-21 (GenBank: MW525061.1; Accession: MW525061). We then uploaded each of the paired-end FASTQ files for each wastewater sample. The Bowtie2 (Galaxy Version 2.5.0) 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 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

Location code	Sample date	VOC (%) ^a	Lineage(s) (%) ^b
VM	10–26-21	Alpha (94.2)	Q.3 (94.2)
VM	11-9-21	Alpha (94.9)	Q.4 (94.7)
VM	9–12-22	Alpha (97.3)	Q.4 (96.9)
VM	3–13-23	Alpha (98.0)	Q.4 (65.5)
VM	4–24-23	Alpha (97.9)	Q.4 (89.7)
VM	5-1-23	Alpha (97.4)	Q.4 (97.1)

Table 1. GT molecular variant calling. ^aRelative abundance of variants of concern (VOC) as a percentage, ^bRelative abundance of VOC lineages as a percentage.

reconstructed and then aligned with the reference gene. Mutations that were present in wastewater samples but not the reference clinical sample were characterized as novel mutations. The total number of reads that aligned to each reference gene were determined in MEGA and FastQC was used to quantify read length and the number of poor-quality sequences (Supplementary Table 1). At least 3 reads were present for each amino acid¹⁵. The same bioinformatics pipeline was used to characterize novel mutations that were present on a curated dataset of Alpha variant sequencing reads that contained lineage-defining mutations.

Novel mutation hotspot analyses

We identified novel mutations as described above. We then tracked the percent prevalence of novel 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 genes. We then uploaded each of the paired-end FASTQ files for each wastewater sample. The Bowtie2 (Galaxy Version 2.5.0) 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 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 mutations. This was repeated for each novel mutation across all samples positive for Alpha variant lineage and the percent prevalence data was represented in a heatmap. Novel mutations were mapped onto 2-D representations of proteins and the 3-D protein structures when available using UCSF Chimera^{15,25}.

Searching GenBank, sequence read archive (SRA), and GISAID online records by mutation

The NCBI Virus SARS-CoV-2 Variants Overview “Search GenBank + SRA Data by Mutation” tool was used to identify online records of each mutation. The total records found include those uploaded to SRA, GenBank, and in unique samples. Similarly, the GISAID SARS-CoV-2 Lineage Comparison Mutation Tracker was used to identify the total number of sequences carrying each mutation and the worldwide prevalence of these mutations in the GISAID database.

Results

Chronic shedding of an alpha variant lineage at a rural WWTP

Wastewater samples were collected between July 2021 and June 2023 and SARS-CoV-2 genome copies per 100 mL wastewater were determined each week and reported to MDHHS^{13,15}. 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. 1A)¹⁵. The increasing level of virus shedding from VM inversely correlated ($r = -0.4506$) with the weekly case count in the associated county during this time course, and during the last two sample collections there were 0 and 1 probable or confirmed COVID-19 cases at the zip code level (Fig. 1B). In order to identify the SARS-CoV-2 variant(s) responsible for this activity, RNA extracted from stored wastewater concentrates was shipped to GT Molecular (Fort Collins, CO) and a next generation sequencing (NGS) and variant calling pipeline was employed. 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)¹⁵. This corresponded to the beginning of the Delta variant wave in Central Michigan¹³. The site of interest began shedding the Alpha variant lineage during the next two time points tested (10–26–21 and 11–9–21). This was preceded by sequencing data from clinical samples, which

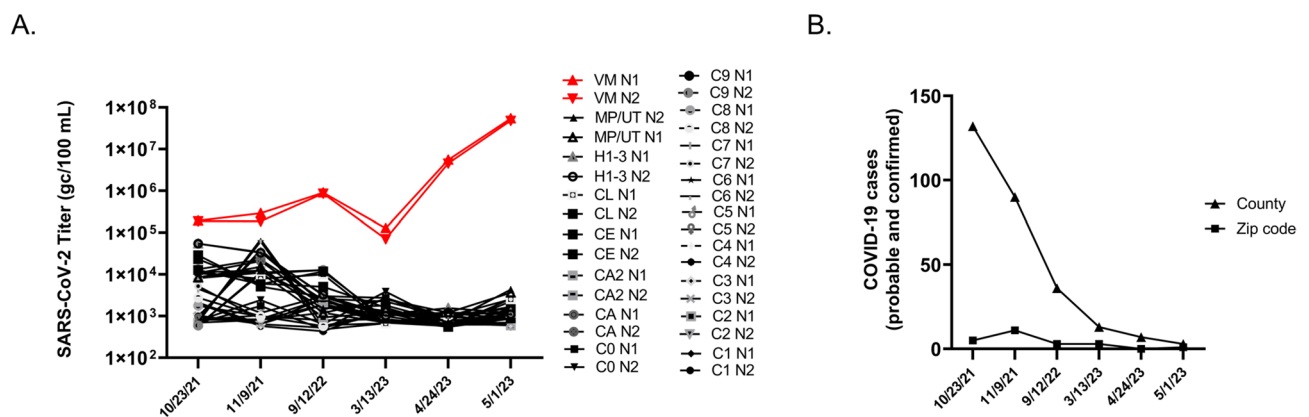


Fig. 1. (A) SARS-CoV-2 genome copies (GC)/100 mL wastewater detected at each weekly sample site from July 2021 to June 2023. Two letter site codes and dates are shown that correspond to sequenced samples. The colors and shapes associated with each sample are located in the graphical legend. VM samples are highlighted in red. (B) Total weekly COVID-19 case counts (probable and confirmed) in VM's county (triangles) and zip code (squares) as reported by the MDHHS Michigan COVID-19 Cases and Deaths dashboard (<https://www.michigan.gov/coronavirus>) and the Central Michigan District Health Department (CMDHD).

revealed 16 Alpha variant lineage Q.3 isolates collected from 2–18–21 to 7–9–21¹⁵. 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)¹⁵. This corresponded to the end of the first Omicron wave in Central Michigan¹³. 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%¹⁵. Specifically, for the samples analyzed in the current manuscript (i.e., 10–26–2021, 11–9–2021, 9–12–2022, 3–13–23, 4–24–23, and 5–1–23), the alpha variant represented 94.2–98% of amplicons. Based on this information, it was possible to reconstruct alpha variant genomes without significant contamination with other variants (Supplementary Data 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 during the same sampling period¹⁵.

Accumulation of novel mutations

We reasoned that chronic shedding of SARS-CoV-2 would lead to accumulation of novel mutations that do not align with sequences identified in most clinical and wastewater samples, which mostly result from acute infection. This hypothesis was supported by our previous analysis of Spike, which identified 37 novel mutations that accumulated during this time frame¹⁵. Alignment of reconstructed consensus genes with a reference Alpha variant lineage clinical sequence revealed that non-Spike proteins had 35 novel mutations in the 5–1–2023 sample including mutations in nsp1¹, nsp2¹, nsp3⁸, nsp4⁴, nsp5/3 CLpro², nsp6², RdRp², nsp15², nsp16¹, ORF3a⁴, M², ORF6¹, ORF7a¹, ORF7b¹, and N³ (Fig. 2). One mutation in both nsp5/3 CLpro and N were reversions to the GISAID reference sequence and therefore weren't investigated further. Each mutation was analyzed using the NCBI Virus SARS-CoV-2 Variant Overview “Search GenBank + SRA Data by Mutation” tool. This analysis identified the total records for each mutation in GenBank and SRA databases during the time frame of Dec 2023 to Jun 2024. Importantly, this time frame was 7 to 13 months after the last wastewater sample was acquired. Very few records were identified for each of these mutations and some mutations lacked records altogether (Tables 2 and 3). We expanded the mutational analysis by quantifying the percent prevalence of each of the 33 novel mutations identified in the 5–1–2023 sample across all wastewater samples that were positive for the Alpha variant lineage. A heatmap showed that these mutations accumulated and became dominant within the population over time, while also retaining diversity at each position (Fig. 3). We repeated this analysis using curated datasets of Alpha variant-specific reads, which overlapped with 3 of the novel mutations. Although this follow-up analysis limited the number of novel mutations that could be investigated, the same pattern revealing an increase in percent prevalence over time was shown (Fig. 4). The GT Molecular variant calling pipeline and follow-up analysis using a curated dataset of Alpha variant-specific sequencing reads strongly support that the novel mutations that we tracked accumulated during a persistent Alpha variant infection.

Nsp1

One novel mutation accumulated in nsp1: S17I. S17I resides in the N-terminal domain (NTD) (Fig. 5A). It was found in 1 online record and was present in <0.5% of all sequences in the GISAID worldwide database (Tables 2 and 3). The impact of this mutation on protein function and immune evasion is unknown.

Nsp2

One novel mutation accumulated in nsp2: G445 C (G265 C) (Fig. 5B). ORF1ab mutations outside of parentheses were determined from the first amino acid in nsp1. ORF1ab mutations inside of parentheses were determined from the first amino acid of each protein. Both variations were searched in the literature. G445 C (G265 C) was reported in a published literature using samples from India and found in a total of 26 online records, and was present in <0.5% of all sequences in the GISAID worldwide database (Tables 2 and 3)²⁶. The impact of this mutation on protein function and immune evasion is unknown.

Nsp3

Eight novel mutations accumulated in nsp3: S944L (S126L), D1184E (D366E), A1306S (A488S), H1545Y (H727Y), K1795Q (K977Q), R2115I (R1297I), H2520 N (H1702 N), and S2661 F (S1843 F). S944L (S126L)

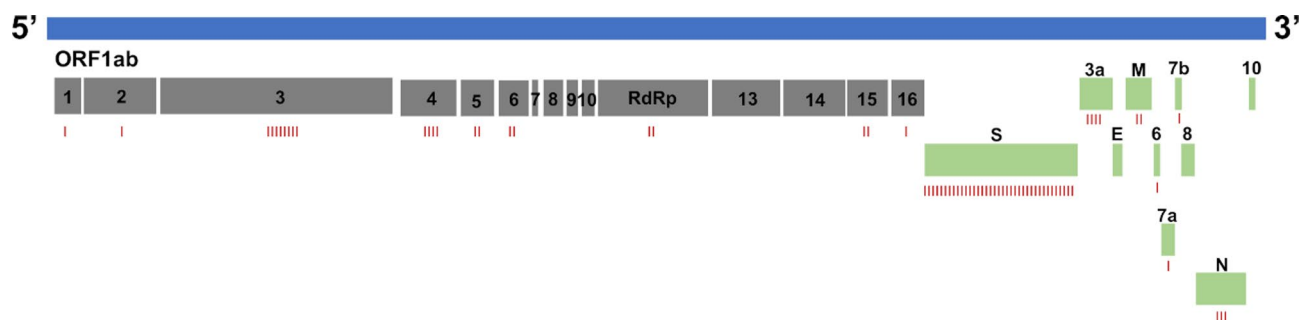


Fig. 2. SARS-CoV-2 genome map with novel mutations identified as red dashes under each open reading frame (ORF): nsp1¹, nsp2¹, nsp3⁸, nsp4⁴, nsp5/3 CLpro², nsp6², RdRp², nsp15², nsp16¹, S³⁷, ORF3a⁴, M², ORF6¹, ORF7a¹, ORF7b¹, and N³.

MUTATION	SRA	GenBank	Unique Samples
NSP1:S17I	1	0	1
NSP2:G445 C (G265 C) ^a	6	20	21
NSP3:S944L (S126L)	60	148	111
NSP3:D1184E (D366E)	6	8	6
NSP3:A1306S (A488S)	0	0	0
NSP3:H1545Y (H727Y)	1	4	2
NSP3:K1795Q (K977Q)	2	11	8
NSP3:R2115I (R1297I)	13	1	13
NSP3:H2520 N (H1702 N)	0	0	0
NSP3:S2661 F (S1843 F)	19	23	25
NSP4:M2796I (M33I)	33	39	41
NSP4:V2943 F (V180 F)	1	2	1
NSP4:I2961 F (I198 F)	0	0	0
NSP4:D2980G (D217G)	11	10	11
3 CLpro: K3499R (K236R)	1	4	4
NSP6:L3606 F (L37 F)	714	984	977
NSP6:F3753 V (F187 V)	2	6	6
RdRp: S4621 N (S232 N)	13	29	18
RdRp: T5304 N (T915 N)	0	0	0
NSP15:E6712 K (E263 K)	2	2	2
NSP15:F6715L (F266L)	6	8	8
NSP16:R7014 N (R219 N)	0	0	0
ORF3a: I35 K	42	56	46
ORF3a: E102D	1	2	2
ORF3a: G172R	0	1	0
ORF3a: M260 K	11	17	15
M: D3 N	13	14	19
M: Q19H	0	0	0
ORF6:D61E	0	0	0
ORF7a: T120I	34	72	63
ORF7b: A43 V	43	68	65
N: N8D	2	3	2
N: S37P	28	48	48

Table 2. Total records of Non-Spike mutations identified in a chronic SARS-CoV-2 alpha variant. ^aORF1ab mutations outside of parentheses were determined from the first amino acid in nsp1. ORF1ab mutations inside of parentheses were determined from the first amino acid of each protein and were required for this analysis.

resides in the hypervariable region (HVR), D1184E (D366E) resides in macrodomain II (Mac2), A1306S (A488S) resides in macrodomain III (Mac3), H1545Y (H727Y) resides in the ubiquitin-like domain 2 (Ubl2), K1795Q (K977Q) resides in the PL2^{pro} domain, R2115I (R1297I) resides within the betacoronavirus-specific marker (βSM), and H2520 N (H1702 N) and S2661 F.

(S1843 F) reside in the Y1 and CoV-Y domains (Fig. 5C). S944L (S126L), A1306S (A488S), and S2661 F (S1843 F) were reported in published literature using Asian isolates and found in a total of 208, 0, and 42 online records, respectively (Table 2)^{27–30}. S944L (S126L), A1306S (A488S), and S2661 F (S1843 F) were present in <0.5, 26, and <0.5% of all sequences in the GISAID worldwide database, respectively (Table 3). H1545Y (H727Y) was reported in published literature using samples from Italy and found in a total of 5 online records, and was present in <0.5% of all sequences in the GISAID worldwide database (Tables 2 and 3)³¹. K1795Q (K977Q) was reported in published literature using samples from Brazil, was present at a higher frequency in immunodeficient patients, and found in a total of 13 online records, and was present in 1% of all sequences in the GISAID worldwide database (Tables 2 and 3)^{32,33}. R2115I (R1297I) was reported in published literature using samples from Moldova and found in a total of 14 online records, and was present in <0.5% of all sequences in the GISAID worldwide database (Tables 2 and 3)³⁴. D1184E (D366E) and H2520 N (H1702 N) have not been previously reported in published literature and were found in a total of 14 and 0 online records, respectively (Table 2). D1184E (D366E) and H2520 N (H1702 N) were both present in <0.5% of all sequences in the GISAID worldwide database (Table 3). The impact of these mutations on nsp3 protein function and immune evasion is unknown.

Nsp4

Four novel mutations accumulated in nsp4: M2796I (M33I), V2943 F (V180 F), I2961 F (I198 F), and D2980G (D217G). M2796I (M33I) resides in a transmembrane domain (TD), and V2943 F (V180 F), I2961 F (I198 F),

Mutation	Total sequences	% prevalence
NSP1:S17I	1,253	< 0.5
NSP2:G445 C (G265 C) ^a	1,400	< 0.5
NSP3:S944L (S126L)	68,199	< 0.5
NSP3:D1184E (D366E)	2,410	< 0.5
NSP3:A1306S (A488S)	3,964,945	26
NSP3:H1545Y (H727Y)	5,255	< 0.5
NSP3:K1795Q (K977Q)	125,281	1
NSP3:R2115I (R1297I)	2,101	< 0.5
NSP3:H2520 N (H1702 N)	70	< 0.5
NSP3:S2661 F (S1843 F)	10,100	< 0.5
NSP4:M2796I (M33I)	11,589	< 0.5
NSP4:V2943 F (V180 F)	2,136	< 0.5
NSP4:I2961 F (I198 F)	11	< 0.5
NSP4:D2980G (D217G)	10,632	< 0.5
3 CLpro: K3499R (K236R)	0	0
NSP6:L3606 F (L37 F)	255,933	2
NSP6:F3753 V (F187 V)	0	0
RdRp: S4621 N (S232 N)	0	0
RdRp: T5304 N (T915 N)	0	0
NSP15:E6712 K (E263 K)	208	< 0.5
NSP15:F6715L (F266L)	0	0
NSP16:R7014 N (R219 N)	0	0
ORF3a: I35 K	8,614	< 0.5
ORF3a: E102D	438	< 0.5
ORF3a: G172R	16,578	< 0.5
ORF3a: M260 K	5,764	< 0.5
M: D3 N	2,049,320	13
M: Q19H	275	< 0.5
ORF6:D61E	121	< 0.5
ORF7a: T120I	4,237,260	27
ORF7b: A43 V	10,155	< 0.5
N: N8D	6,456	< 0.5
N: S37P	9,756	< 0.5

Table 3. GISAID lineage mutation tracker analysis. ^aORF1ab mutations outside of parentheses were determined from the first amino acid in nsp1. ORF1ab mutations inside of parentheses were determined from the first amino acid of each protein. The GISAID database that was queried was updated as of 5 March 2025.

and D2980G (D217G) reside in the luminal domain (LD) (Fig. 5D). M2796I (M33I) was reported in published literature using samples from the Middle East and computational analysis suggested that the mutation causes secondary structure changes converting an alpha helix to a beta sheet, which may impact interaction between nsp3 and nsp4^{35,36}. D2980G (D217G) and M2796I (M33I) were reported in published literature using Asian isolates and were found in a total of 21 and 72 online records, respectively (Table 2)^{37,38}. D2980G (D217G) and M2796I (M33I) were both present in <0.5% of all sequences in the GISAID worldwide database (Table 3). V2943 F (V180 F) and I2961 F (I198 F) have not been previously reported in published literature and were found in a total of 3 and 0 online records, respectively (Table 2). V2943 F (V180 F) and I2961 F (I198 F) were both present in <0.5% of all sequences in the GISAID worldwide database (Table 3). The impact of these mutations on protein function and immune evasion are unknown.

Nsp5/3 CLpro

One novel mutation accumulated in nsp5: K3499R (K236R). K3499R (K236R) resides in D3 (Fig. 5E). K3499R (K236R) was reported in published literature using samples from Indian isolates and was found in a total of 5 online records, and was not present in any sequences in the GISAID worldwide database (Tables 2 and 3)³⁷. The impact of this mutation on protein function and immune evasion is unknown.

Nsp6

Two novel mutations accumulated in nsp6: L3606 F (L37 F) and F3753 V (F187 V). L3606 F (L37 F) resides in a transmembrane domain (TD) and F3753 V (F187 V) resides in a luminal domain (LD) (Fig. 5F). L3606 F (L37 F) was reported in literature using Asian isolates and was found in a total of 1,698 (2%) online records, and was present in 2% of all sequences in the GISAID worldwide database (Tables 2 and 3)³⁹. Previous research

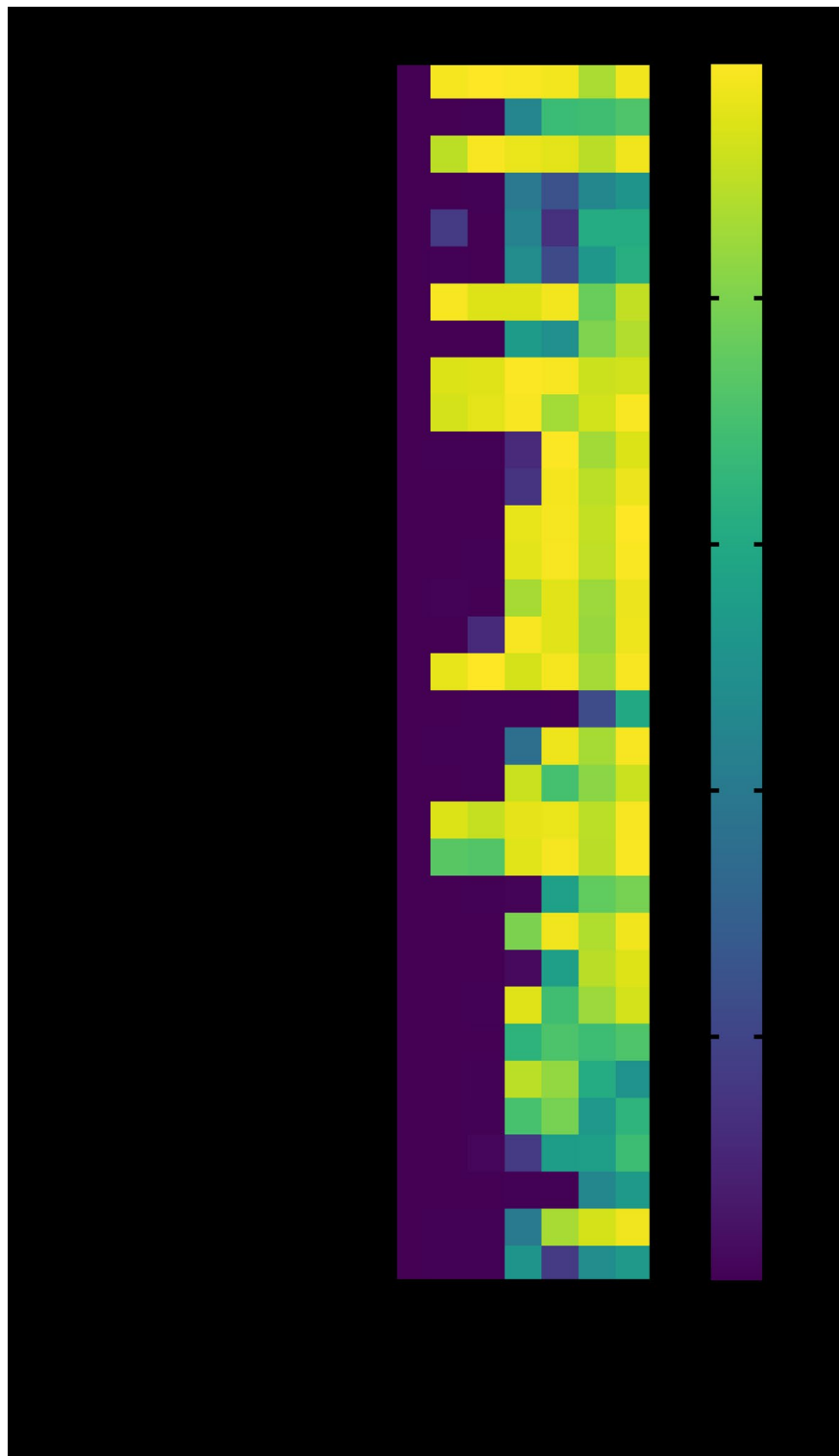


Fig. 3. Heatmap showing the percent prevalence of novel non-Spike mutations in each wastewater sample that was positive for the Alpha variant lineage. ORF1ab mutations outside of parentheses were determined from the first amino acid in nsp1. ORF1ab mutations inside of parentheses were determined from the first amino acid of each protein.

found that L3606 F (L37 F) reduced nsp6's interaction with ATP6 AP1. This allowed for lysosomal acidification to proceed normally, which prevented activation of the NLRP3 inflammasome pathway. The investigators noted that this mutation reduced SARS-CoV-2 fitness, and that this may be why the mutation is not present in circulating variants of concern^{40,41}. F3753 V (F187 V) was reported in a basic science study as an adaptive



Fig. 4. Heatmap showing the percent prevalence of novel non-Spike mutations identified on reads containing Alpha variant lineage-defining mutations. ORF1ab mutations outside of parentheses were determined from the first amino acid in nsp1. ORF1ab mutations inside of parentheses were determined from the first amino acid of each protein.

mutation that arose in ferrets and in vitro during treatment with remdesivir^{16,17}. This mutation was not reported in clinical sequences but was present in a total of 8 online records and was not present in any sequence in the GISAID worldwide database (Tables 2 and 3).

RdRp

Two novel mutations accumulated in RdRp: S4621 N (S232 N) and T5304 N (T915 N). S4621 N (S232 N) resides in the NTD and T5304 N (T915 N) resides in the thumb domain (TD) (Fig. 5G). S4621 N (S232 N) was not previously reported in published literature but was present in a total of 42 online records and was not present in any sequence in the GISAID worldwide database (Tables 2 and 3). T5304 N (T915 N) was reported in published literature using samples from Italy and was present in a total of 0 online records and was not present in any sequence in the GISAID worldwide database (Tables 2 and 3)⁴². The impact of these mutations on protein function and immune evasion are unknown.

Nsp15

Two novel mutations accumulated in nsp15: E67012 K (E263 K) and F6715L (F266L). Both E6712 K (E263 K) and F6715L (F266L) reside in the nuclease EndoU domain (Fig. 5H). E67012 K (E263 K) was reported in literature using samples from Italy and was present in a total of 4 online records and was present in <0.5% of sequences in the GISAID worldwide database (Tables 2 and 3). This mutation was associated with an increased frequency of mortality⁴³. F6715L (F266L) was not previously reported in literature and was present in a total of 14 online records and was not present in any sequence in the GISAID worldwide database (Tables 2 and 3). The impact of these mutations on protein function and immune evasion are unknown.

Nsp16

One novel mutation accumulated in nsp16: R7014 N (R219 N) (Fig. 5I). R7014 N (R219 N) was reported in literature using samples from Brazil and was present in a total of 0 online records and was not present in any sequence in the GISAID worldwide database (Tables 2 and 3). This mutation was associated with reinfection of a healthcare worker¹⁹. The impact of this mutation on protein function and immune evasion is unknown.

ORF3a

Four novel mutations accumulated in ORF3a: I35 K, E102D, G172R, and M260 K. I35 K resides in the N-terminal domain (NTD), E102D resides in the transmembrane domain (TD), and both G172R and M260 K reside in the C-terminal domain (CTD) (Fig. 6A). I35 K and E102D have not been previously reported in literature and were present in <0.5% of sequences found in the GISAID worldwide database (Tables 2 and 3). G172R and M260 K were previously reported in literature analyzing SARS-CoV-2 variants and were present in <0.5% of sequences found in the GISAID worldwide database (Tables 2 and 3)^{44–47}. The impact of these mutations on ORF3a protein function and immune evasion is unknown.

M

Two novel mutations accumulated in M: D3 N and Q19H. Both mutations reside in the N-terminal domain (NTD), which are surface exposed (Fig. 6B). D3 N was previously identified in BA.5 strains and may result in N-myristoylation possibly impacting membrane integrity, post-translational modification, and immune evasion^{48,49}. D3 N was present in 13% of sequences found in the GISAID worldwide database (Table 3). Q19H has not been previously reported in literature and its impact on protein function is unknown. Q19H was present in <0.5% of the sequences found in the GISAID worldwide database (Table 3).

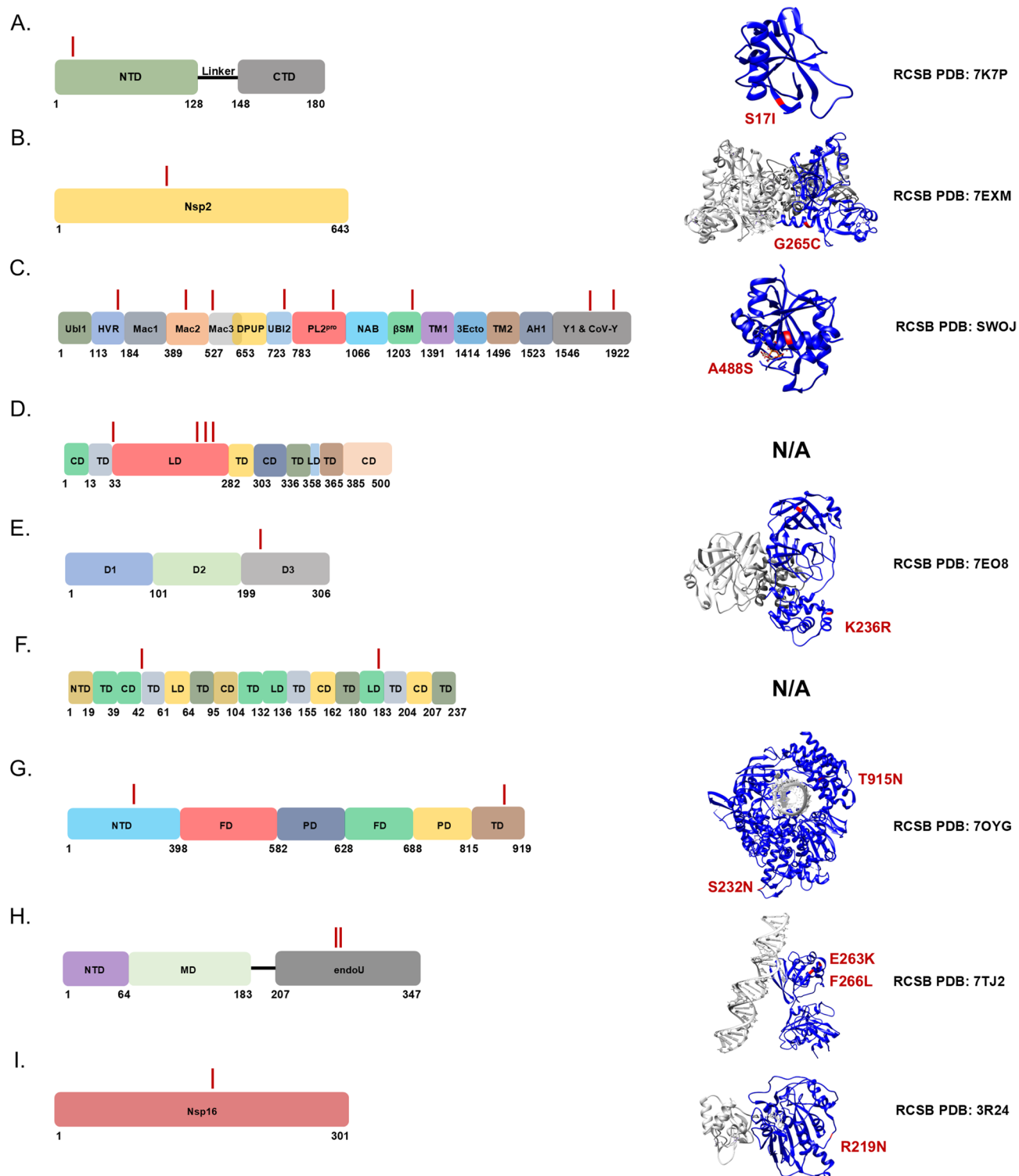


Fig. 5. Novel mutations present in ORF1ab genes were mapped onto 2-D and 3-D protein models: (A) nsp1, (B) nsp2, (C) nsp3, (D) nsp4, (E) nsp5/3 CLpro, (F) nsp6, (G) RdRp, (H) nsp15, and (I) nsp16. Structures were rendered using UCSF Chimera and RCSB PDB numbers were provided⁵⁸. Individual protein domains were indicated in blue and multimers/other molecules were indicated as gray. Mutations were highlighted and noted in red.

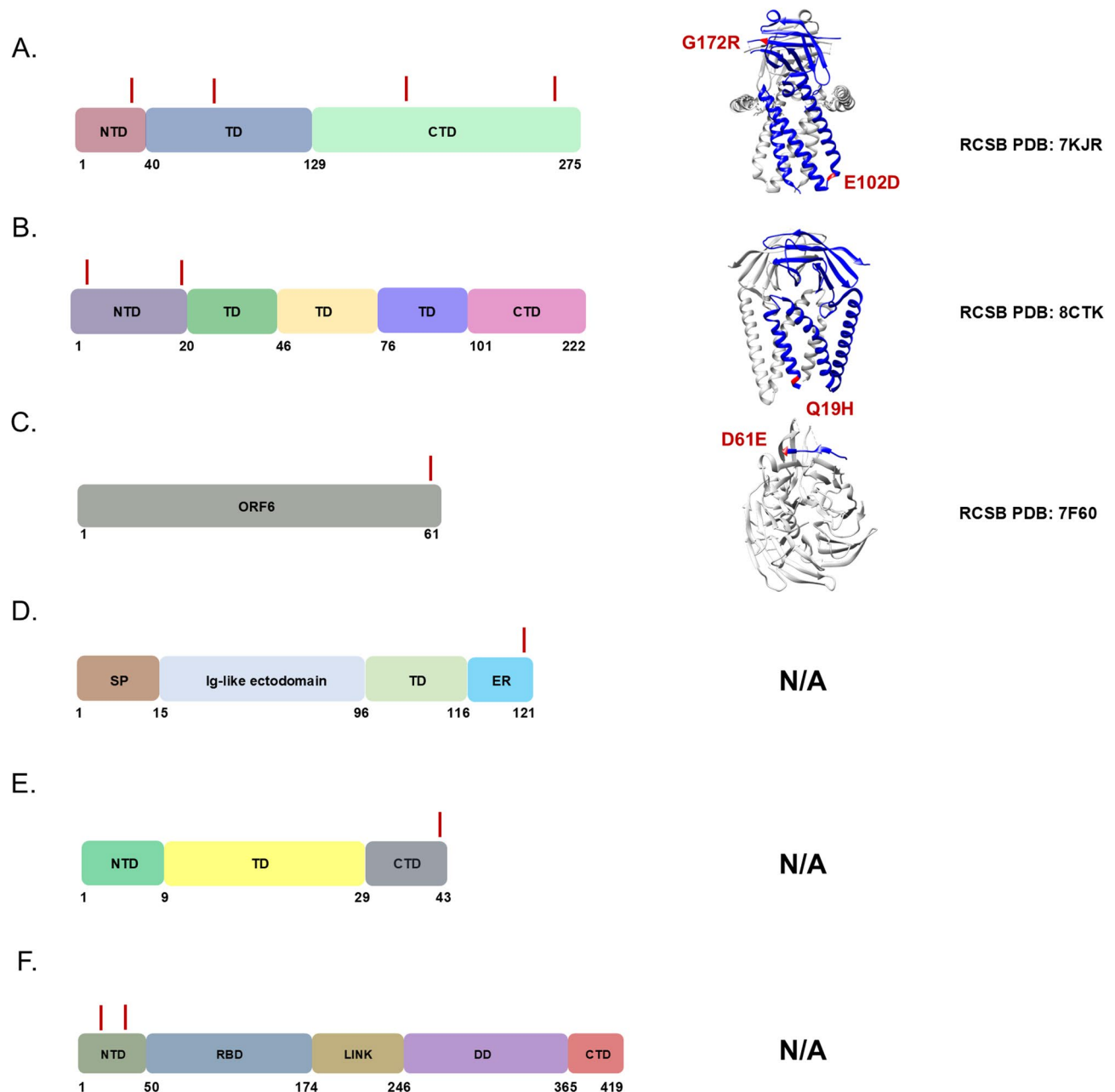


Fig. 6. Novel mutations present in ORF genes were mapped onto 2-D and 3-D protein models. (A) ORF3a, (B) M, (C) ORF6, (D) ORF7a, (E) ORF7b, and (F) N. Structures were rendered using UCSF Chimera and RCSB PDB numbers were provided⁵⁸. Individual protein domains were indicated in blue and multimers/other molecules were indicated as gray. Mutations were highlighted and noted in red.

ORF6

One novel mutation accumulated in ORF6: D61E (Fig. 6C). This is the last residue in the open reading frame. D61E has not been previously reported in literature and was present in < 0.5% of sequences found in the GISAID worldwide database (Tables 2 and 3). The impact of this mutation on ORF6 protein function and immune evasion is unknown.

ORF7a

One novel mutation accumulated in ORF7a: T120I. This is the second to last C-terminal residue and resides in the endoplasmic reticulum retention sequence (Fig. 6D). T120I has been previously identified in a SARS-CoV-2 variant and was present in 27% of sequences found in the GISAID worldwide database⁵⁰. The impact of this mutation on ORF7a protein function and immune evasion is unknown.

ORF7b

One novel mutation accumulated in ORF7b: A43 V. This is the last residue in the C-terminal domain (Fig. 6E). A43 V has not been previously reported in literature and was present in <0.5% of sequences found in the GISAID worldwide database (Tables 2 and 3). The impact of this mutation on ORF7b protein function and immune evasion is unknown.

N

Two novel mutations accumulated in the N gene: N8D and S37P. Both N8D and S37P reside in the N-terminal domain (NTD) (Fig. 6F). N8D was previously identified in B.1.1.7 strains and was present in <0.5% of sequences found in the GISAID worldwide database (Tables 2 and 3)^{51,52}. S37P was previously identified in SARS-CoV-2 isolates and was present in <0.5% of sequences found in the GISAID worldwide database (Tables 2 and 3)^{53,54}. The impact of these mutations on N protein function and immune evasion is unknown.

Discussion

Retrospective analysis of wastewater data revealed that one rural site produced consistently higher concentrations of SARS-CoV-2 copy numbers despite a decreasing COVID-19 case count in the associated county during this time course. 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 and the fact that only 3 COVID-19 cases were present in the entire county in 5-1-2023, our October 2021 Alpha variant lineage reconstruction may represent a chronic infection that lasted for 2–7 months. If this were true, at this stage of the chronic infection, the Alpha variant lineage already accumulated 9 novel mutations in the Spike gene and 10 novel mutations spread across nsp1, nsp3, nsp5, nsp6, nsp15, nsp16, and N¹⁵. It is important to note that we did not confirm that this was due to a chronic infection in a human and a non-human source is possible.

Eighteen months later, the Alpha variant lineage shed from this site had 72 novel non-structural (nsp1, nsp2, nsp3, nsp4, nsp5/3 CLpro, nsp6, RdRp, nsp15, nsp16, ORF3a, ORF6, ORF7a, and ORF7b) and structural (Spike, M, and N) mutations¹⁵. Spike mutations present in these wastewater samples have been described previously. Many of the Spike mutations were associated with experimental evidence suggesting they promoted immune evasion, and three mutations were previously found in immunocompromised patients¹⁵. Similarly, nsp3 K1795Q (K977Q) was present at a higher frequency in immunodeficient patients, nsp6 F3753 V (F187 V) was an adaptive mutation that accumulated in the presence of remdesivir in vitro and in vivo, and nsp16 R7014 N (R219 N) was present in a reinfecting healthcare worker^{16,17,19,33}. Considering that 8% of the novel mutations in Spike and non-Spike genes were associated with known persistent infections, we can assume that many of the identified mutations are selected during chronic infection.

Without experimental evidence, it is impossible to know the precise role of each mutation, although we can infer based on known structural and functional information for each protein, in addition to serological data in the human host. For instance, the two mutations present in M protein (D3 N and Q19H) reside in the N-terminal domain (NTD) and are surface exposed^{55,56}. Previous research has identified D3 N in BA.5 strains and investigators suggested that the mutation may result in N-myristoylation possibly impacting membrane integrity, post-translational modification, and immune evasion^{48,49}. Similar to Spike, antibodies directed to M are generated during infection, persist for at least a year after infection, and generate a similar level of reactivity as immunodominant linear epitopes^{55–57}. It's possible that during a chronic infection SARS-CoV-2 optimizes Spike and M to evade adaptive immunity.

These data provide indirect evidence that an individual can be chronically infected with SARS-CoV-2 over many months and possibly a few years. It is possible that multiple individuals contributed to the persistence of this variant of concern, or perhaps a non-human organism was infected that shed virus into the sewer system. Previous research has revealed the presence of cryptic mutations in wastewater samples that are not prevalent in clinical samples and that may be due to persistent shedding from humans or animals [1–7]. During persistent infection, SARS-CoV-2 can accumulate many mutations in Spike and non-Spike genes. Some of these mutations have been previously reported in persistent infections, although most of them are rare and poorly described, present in less than 0.5% of all sequences found in the GISAID worldwide database. The rarity of these mutations in the GISAID database may reflect the rarity of a chronic infection in humans, or another rare event that led to persistent shedding in this sewer shed. Further research is needed to determine which of these mutations are predictive of chronic infection and if they can be used as a biomarker in individuals with persistent disease and leveraged to tailor selection or development of pharmaceutical therapies. This study also shows that small WWTPs can enhance the resolution of rare biological events and allow for total reconstruction of viral genomes and their corresponding proteins.

Data availability

FASTQ files for each sample are available in the NCBI SRA database (SRA: SRP465974; BioProject ID: PRJ-NA1027333).

Received: 5 November 2024; Accepted: 22 May 2025

Published online: 28 May 2025

References

1. Anderson-Coughlin, B. L. 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.* **830**, 154619 (2022).
2. Betancourt, W. Q. et al. COVID-19 containment on a college campus via wastewater-based epidemiology, targeted clinical testing and an intervention. *Sci. Total Environ.* **779**, 146408 (2021).
3. Corchis-Scott, R. et al. Averting an outbreak of SARS-CoV-2 in a university residence hall through wastewater surveillance. *Microbiol. Spectr.* **9** (2), e0079221 (2021).
4. Gibas, C. et al. Implementing building-level SARS-CoV-2 wastewater surveillance on a university campus. *Sci. Total Environ.* **782**, 146749 (2021).
5. Jarvie, M. et al. Thi. RT-ddPCR Wastewater Monitoring of COVID-19 Across the Eastern Upper Peninsula of Michigan. SSRN2022.
6. Scott, L. C. et al. Targeted wastewater surveillance of SARS-CoV-2 on a university campus for COVID-19 outbreak detection and mitigation. *Environ. Res.* **200**, 111374 (2021).
7. Wong, T. E. et al. Evaluating the sensitivity of SARS-CoV-2 infection rates on college campuses to wastewater surveillance. *Infect. Dis. Model.* **6**, 1144–1158 (2021).
8. Rondeau, N. C. et al. Building-Level detection threshold of SARS-CoV-2 in wastewater. *Microbiol. Spectr.* :e0292922. (2023).
9. Chua, F. J. D. 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.* **875**, 162611 (2023).
10. Vo, V. et al. Detection of the Omicron BA.1 variant of SARS-CoV-2 in wastewater from a Las Vegas tourist area. *JAMA Netw. Open.* **6** (2), e230550 (2023).
11. Pico-Tomás, A. et al. Surveillance of SARS-CoV-2 in sewage from buildings housing residents with different vulnerability levels. *Sci. Total Environ.* **872**, 162116 (2023).
12. Lehto, K. M. et al. Wastewater-based surveillance is an efficient monitoring tool for tracking influenza A in the community. *Water Res.* **257**, 121650 (2024).
13. Conway, M. J. et al. SARS-CoV-2 wastewater monitoring in rural and small metropolitan communities in central Michigan. *Sci. Total Environ.* **894**, 165013 (2023).
14. Galloway, S. E. et al. Emergence of SARS-CoV-2 B.1.1.7 Lineage - United States, December 29, 2020-January 12, 2021. *MMWR Morb Mortal. Wkly. Rep.* **70** (3), 95–99 (2021).
15. Conway, M. J. et al. Chronic shedding of a SARS-CoV-2 alpha variant in wastewater. *BMC Genom.* **25** (1), 59 (2024).
16. Checkmahomed, L. et al. In vitro selection of Remdesivir-Resistant SARS-CoV-2 demonstrates high barrier to resistance. *Antimicrob. Agents Chemother.* **66** (7), e0019822 (2022).
17. Cox, R. M. et al. Oral prodrug of Remdesivir parent GS-441524 is efficacious against SARS-CoV-2 in ferrets. *Nat. Commun.* **12** (1), 6415 (2021).
18. Kemp, S. A. et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **592** (7853), 277–282 (2021).
19. Camargo, C. H. et al. SARS-CoV-2 reinfection in a healthcare professional in inner Sao Paulo during the first wave of COVID-19 in Brazil. *Diagn. Microbiol. Infect. Dis.* **101** (4), 115516 (2021).
20. Ahmadi, A. S. et al. SARS-CoV-2 intrahost evolution in immunocompromised patients in comparison with immunocompetent populations after treatment. *J. Med. Virol.* **95** (6), e28877 (2023).
21. Flood, M. T., D'Souza, N., Rose, J. B. & Aw, T. G. Methods evaluation for rapid concentration and quantification of SARS-CoV-2 in Raw wastewater using droplet digital and quantitative RT-PCR. *Food Environ. Virol.* **13** (3), 303–315 (2021).
22. Lu, X. et al. US CDC Real-Time reverse transcription PCR panel for detection of severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis.* **26** (8), 1654–1665 (2020).
23. Lee, H. W., Lee, H. M., Yoon, S. R., Kim, S. H. & Ha, J. H. Pretreatment with Propidium monoazide/sodium Lauroyl sarcosinate improves discrimination of infectious waterborne virus by RT-qPCR combined with magnetic separation. *Environ. Pollut.* **233**, 306–314 (2018).
24. Karthikeyan, S. et al. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature* **609** (7925), 101–108 (2022).
25. Wrobel, A. G. et al. SARS-CoV-2 and Bat RaTG13 Spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat. Struct. Mol. Biol.* **27** (8), 763–767 (2020).
26. Sarkar, R. et al. Emergence of a novel SARS-CoV-2 Pango lineage B.1.1.526 in West Bengal, India. *J. Infect. Public. Health.* **15** (1), 42–50 (2022).
27. Shah, A. et al. Comparative mutational analysis of SARS-CoV-2 isolates from Pakistan and structural-functional implications using computational modelling and simulation approaches. *Comput. Biol. Med.* **141**, 105170 (2022).
28. Abdullaev, A. et al. Genome sequence diversity of SARS-CoV-2 obtained from clinical samples in Uzbekistan. *PLoS One.* **17** (6), e0270314 (2022).
29. Zhu, M. et al. Tracking the molecular evolution and transmission patterns of SARS-CoV-2 lineage B.1.466.2 in Indonesia based on genomic surveillance data. *Viol. J.* **19** (1), 103 (2022).
30. Laskar, R. & Ali, S. Mutational analysis and assessment of its impact on proteins of SARS-CoV-2 genomes from India. *Gene* **778**, 145470 (2021).
31. De Marco, C. et al. Whole-genome analysis of SARS-CoV-2 in a 2020 infection cluster in a nursing home of Southern Italy. *Infect. Genet. Evol.* **99**, 105253 (2022).
32. Ramesh, S. et al. Emerging SARS-CoV-2 variants: A review of its mutations, its implications and vaccine efficacy. *Vaccines (Basel)* **9**(10). (2021).
33. Wilkinson, S. A. J. et al. Recurrent SARS-CoV-2 mutations in immunodeficient patients. *Virus Evol.* **8** (2), veac050 (2022).
34. Ulinici, M. et al. Genome sequences of SARS-CoV-2 strains from the Republic of Moldova. *Microbiol. Resour. Announc.* **12** (1), e0113222 (2023).
35. Rehman, S., Mahmood, T., Aziz, E. & Batool, R. Identification of novel mutations in SARS-COV-2 isolates from Turkey. *Arch. Virol.* **165** (12), 2937–2944 (2020).
36. Abou-Hamdan, M. et al. Variant analysis of the first Lebanese SARS-CoV-2 isolates. *Genomics* **113** (1 Pt 2), 892–895 (2021).
37. Das, J. K., Sengupta, A., Choudhury, P. P. & Roy, S. Characterizing genomic variants and mutations in SARS-CoV-2 proteins from Indian isolates. *Gene Rep.* **25**, 101044 (2021).
38. Sassi, M. B. et al. Phylogenetic and amino acid signature analysis of the SARS-CoV-2s lineages Circulating in Tunisia. *Infect. Genet. Evol.* **102**, 105300 (2022).
39. Omotoso, O. E., Olugbami, J. O. & Gbadegesin, M. A. Assessment of intercontinents mutation hotspots and conserved domains within SARS-CoV-2 genome. *Infect. Genet. Evol.* **96**, 105097 (2021).
40. Bills, C., Xie, X. & Shi, P. Y. The multiple roles of nsp6 in the molecular pathogenesis of SARS-CoV-2. *Antiviral Res.* **213**, 105590 (2023).
41. Sun, X. et al. SARS-CoV-2 non-structural protein 6 triggers NLRP3-dependent pyroptosis by targeting ATP6AP1. *Cell. Death Differ.* **29** (6), 1240–1254 (2022).
42. Equestre, M. et al. Characterization of SARS-CoV-2 variants in military and civilian personnel of an air force airport during three pandemic waves in Italy. *Microorganisms* **11**(11). (2023).
43. Fang, S. et al. Updated SARS-CoV-2 single nucleotide variants and mortality association. *J. Med. Virol.* **93** (12), 6525–6534 (2021).

44. Tierling, S. et al. Rapid Base-Specific calling of SARS-CoV-2 variants of concern using combined RT-PCR melting curve screening and SIRPH technology. *Open. Forum Infect. Dis.* **8** (8), ofab364 (2021).
45. Patro, L. P. P., Sathyaseelan, C., Uttamrao, P. P. & Rathinavelan, T. The evolving proteome of SARS-CoV-2 predominantly uses mutation combination strategy for survival. *Comput. Struct. Biotechnol. J.* **19**, 3864–3875 (2021).
46. Azad, G. K. & Khan, P. K. Variations in Orf3a protein of SARS-CoV-2 alter its structure and function. *Biochem. Biophys. Rep.* **26**, 100933 (2021).
47. Franceschi, V. B. et al. Predominance of the SARS-CoV-2 lineage P.1 and its sublineage P.1.2 in patients from the metropolitan region of Porto Alegre, Southern Brazil in March 2021. *Pathogens* **10**(8). (2021).
48. Hossain, A., Akter, S., Rashid, A. A., Khair, S. & Alam, A. Unique mutations in SARS-CoV-2 Omicron subvariants' non-spike proteins: potential impacts on viral pathogenesis and host immune evasion. *Microb. Pathog.* **170**, 105699 (2022).
49. Tegally, H. et al. Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. *Nat. Med.* **28** (9), 1785–1790 (2022).
50. Cruz, C. A. K. & Medina, P. M. B. Temporal changes in the accessory protein mutations of SARS-CoV-2 variants and their predicted structural and functional effects. *J. Med. Virol.* **94** (11), 5189–5200 (2022).
51. Zárate, S. et al. The alpha variant (B.1.1.7) of SARS-CoV-2 failed to become dominant in Mexico. *Microbiol. Spectr.* **10** (2), e0224021 (2022).
52. Dimonte, S., Babakir-Mina, M., Hama-Soor, T. & Ali, S. Genetic variation and evolution of the 2019 novel coronavirus. *Public. Health Genomics.* **24** (1–2), 54–66 (2021).
53. Mohammad, T. et al. Genomic variations in the structural proteins of SARS-CoV-2 and their deleterious impact on pathogenesis: A comparative genomics approach. *Front. Cell. Infect. Microbiol.* **11**, 765039 (2021).
54. Timmers, L. et al. SARS-CoV-2 mutations in Brazil: from genomics to putative clinical conditions. *Sci. Rep.* **11** (1), 11998 (2021).
55. Williams, D. M. et al. Establishing SARS-CoV-2 membrane protein-specific antibodies as a valuable serological target via high-content microscopy. *iScience* **26** (7), 107056 (2023).
56. Jörrißen, P. et al. Antibody response to SARS-CoV-2 membrane protein in patients of the acute and convalescent phase of COVID-19. *Front. Immunol.* **12**, 679841 (2021).
57. Sullivan, D. J., Franchini, M., Joyner, M. J., Casadevall, A. & Focosi, D. Analysis of anti-SARS-CoV-2 Omicron-neutralizing antibody titers in different vaccinated and unvaccinated convalescent plasma sources. *Nat. Commun.* **13** (1), 6478 (2022).
58. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25** (13), 1605–1612 (2004).

Acknowledgements

We thank MDHHS and MiNET for supporting wastewater collection, processing, sequencing, and data analysis and CMDHD (Steve King, Taylor Irwin, Norman Kean) for access to local and regional case data. We also thank the WWTP staff who provided samples every week during a pandemic. We thank CMU undergraduate assistants Justus Holben, Gabrielle Reau, Jessica Broach, Hamzah Khan, Jayde-Ann Taylor, Ashley Bergmooser, Kaitlyn Perry, Emily Rosema, and Alexis Bruce for their support, and Kristopher J. Kieft for bioinformatics expertise. This is contribution number 215 of the Central Michigan University Institute for Great Lakes Research.

Author contributions

MJC wrote and revised the manuscript and directed the wastewater monitoring activities, MPN reconstructed ORFs, CMP 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.

Funding

Michigan Department of Health and Human Services (MDHHS).

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-03771-5>.

Correspondence and requests for materials should be addressed to M.J.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025