



OPEN

## Local monitoring of SARS-CoV-2 variants in two large California counties in 2021

Noah Kojima<sup>1✉</sup>, Eugenia Khorosheva<sup>2</sup>, Lauren Lopez<sup>2</sup>, Mikhail Hanewich-Hollatz<sup>2</sup>, J. Cesar Ignacio-Espinoza<sup>2</sup>, Matthew Brobeck<sup>2</sup>, Janet Chen<sup>2</sup>, Matthew Geluz<sup>2</sup>, Victoria Hess<sup>2</sup>, Sophia Quasem<sup>2</sup>, Nabjot Sandhu<sup>2</sup>, Elias Salfati<sup>2</sup>, Maria Shacreaw<sup>2</sup>, George Way<sup>2</sup>, Zhiyi Xie<sup>2</sup>, Vladimir Slepnev<sup>2</sup> & Jeffrey D. Klausner<sup>3</sup>

Coronavirus Disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), continues to persist due to mutations resulting in newer, more infectious variants of concern. We aimed to leverage an ongoing private SARS-CoV-2 testing laboratory's infrastructure to monitor SARS-CoV-2 variants in two large California counties. Study enrollment was offered to adults aged 18 years or older in Los Angeles County and Riverside County who recently tested positive for SARS-CoV-2 with a polymerase chain reaction (PCR) assay. A cycle threshold value less than or equal to 30 cycles was considered a positive test for sequencing purposes. Within 5 days of study enrollment, clinician-monitored, self-collected oral fluid and anterior nares swab specimens were obtained from participants. Specimens were transported and stored at 8 °C or cooler. Samples underwent ribonucleic acid extraction, library preparation, and sequencing. SARS-CoV-2 lineages were identified using sequencing data. Participant and genomic data were analyzed using statistical tools and visualized with toolkits. The study was approved by Advarra Institutional Review Board (Pro00053729). From May 27, 2021 to September 9, 2021, 503 individuals were enrolled and underwent specimen collection. Of the 503 participants, 238 (47.3%) participants were women, 329 (63.6%) participants were vaccinated, and 221 (43.9%) participants were of Hispanic or Spanish origin. Of the cohort, 496 (98.6%) participants had symptoms at the time of collection. Among the 503 samples, 443 (88.1%) nasal specimens and 353 (70.2%) oral specimens yielded positive sequencing results. Over our study period, the prevalence of the Alpha variant of SARS-CoV-2 decreased (initially 23.1% [95% confidence interval (95% CI): 0–0.49%] to 0% [95% CI 0.0–0.0%]) as the prevalence of the Delta variant of SARS-CoV-2 increased (initially 33.3% [95% CI 0.0–100.0%] to 100.0% [95% CI 100.0–100.0%]). A strain that carried mutations of both Delta and Mu was identified. We found that outpatient SARS-CoV-2 variant surveillance could be conducted in a timely and accurate manner. The prevalence of different variants changed over time. A higher proportion of nasal specimens yielded results versus oral specimens. Timely and regional outpatient SARS-CoV-2 variant surveillance could be used for public health efforts to identify changes in SARS-CoV-2 strain epidemiology.

Coronavirus disease 2019 (COVID-19) cases, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), continue to persist. Like many viruses, SARS-CoV-2 mutates<sup>1</sup>. When those mutations occur in key parts of important proteins, there is the potential for the virus to become more infectious, resistant to treatments, and able to avoid immunity that could be due to vaccination or natural infection<sup>2,3</sup>.

Following widespread infections caused by the wildtype strain of SARS-CoV-2, the first identified variant of concern was the Alpha variant (UK variant), which spread approximately 50% faster than the wildtype strain<sup>2</sup>. Following the Alpha variant, the Beta (South Africa) and Gamma (Brazilian) variants of concern were identified<sup>4</sup>. The Delta variant of SARS-CoV-2 followed becoming the dominant variant of concern in the United States<sup>5</sup>, with new variants actively being identified<sup>6</sup>. Compared to the wildtype strain of SARS-CoV-2, the Delta variant is more infectious and has the potential to escape vaccination-mediated immune protection<sup>2,7–9</sup>.

<sup>1</sup>Department of Medicine, University of California Los Angeles, 10833 Le Conte Ave, Los Angeles, CA 90095, USA. <sup>2</sup>Curative Inc., San Dimas, CA, USA. <sup>3</sup>Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles 90033, USA. ✉email: nkojima@ucla.edu

Due to the major public health concern of SARS-CoV-2, surveillance programs are needed to monitor SARS-CoV-2 variants to inform public health efforts. With the development of new sequencing tools, it is possible for genomic data to be used to inform those public health responses as novel mutations are identified in widely circulating strains of SARS-CoV-2. The World Health Organization proposed that a strong and resilient global sequencing network that provides useable and timely results is needed to maximize the public health impact of sequencing<sup>10</sup>. Regarding that important effort, we aimed to leverage an ongoing private SARS-CoV-2 testing laboratory's infrastructure to monitor SARS-CoV-2 variants in two large counties in California.

## Methods

**Enrollment.** Study enrollment was offered to eligible adults (18 years or older) in Los Angeles County and Riverside County who tested positive for SARS-CoV-2 by a reverse transcription polymerase chain reaction (RT-PCR) assay conducted at a private laboratory (Curative, San Dimas, CA). To meet eligibility criteria, potential participants were required to have a positive SARS-CoV-2 RT-PCR result with a cycle threshold value less than or equal to 30 cycles within 5 days of study enrollment and time of sample collection for sequencing. Vulnerable subjects, i.e., pregnant women, nursing home residents or other institutionalized people, prisoners, and persons without decisional capacity, were excluded from the study. Written informed consent was obtained from each participant prior to study enrollment. Demographic information and vaccination status was collected from participants with a survey.

**Specimen collection and transport.** Trained healthcare workers directly observed participants as they self-collecting oral fluid and anterior nares swab specimens. Specimens placed into 10 mL collection tubes containing DNA/RNA Shield Stabilization Solution (Zymo Research, Irvine, CA). Samples were transported to the laboratory at 2–8 °C within 5 h of specimen collection and stored at 4 °C for up to 7 days before library preparation.

**Ribonucleic acid (RNA) isolation.** RNA extraction was performed using silica-based. Either manual silica column-based extraction (Total RNA Purification Kit 96 Deep Well Plate Format Dx, Norgen Biotek Corp., Thorold, ON) or a modified automated magnetic silica beads-based extraction method (Thermo Scientific King-Fisher™ Flex, Thermo Scientific, Waltham, MA) were used for extraction. RNA was eluted in 60 µL of 10 mM Tris (pH 7.4).

**Library preparation and sequencing.** The libraries were prepared using an Illumina COVIDSeq protocol on 96-well plates (Illumina, Inc., San Diego, CA). Samples were processed alongside a positive control (SARS-CoV-2 BEI NR-52287 genotype A) and a negative control of human nasal specimen without SARS-CoV-2 RNA. An indexed sample library from each plate was pooled and quantified using a fluorometer (Qubit 3.0, Invitrogen, Waltham, MA). Four 96-well plates were combined at equimolar concentrations and sequenced. Dilution and loading were performed per the manufacturer instruction (Illumina, Inc., San Diego, CA). Dual-indexed paired-end sequencing was performed to get a deeper sequencing depth. Sequencing aimed to have 1–2 million reads per sample (NextSeq2000, Illumina, Inc., San Diego, CA).

**Quality control of reads.** Paired-end reads were filtered and trimmed to eliminate the presence of primers and adapters. Data quality was assessed to ensure high-quality reads. A minimum quality score of 30 was selected (*pTrimer-1.3.4 -a Primers.bed -q 30 -t pair*)<sup>11</sup>.

**Mapping.** Paired-end reads were then mapped against the ‘Wuhan seafood market pneumonia virus isolate’ Wuhan-Hu-1 genome (*Accession number: NC\_045512.2*)<sup>12</sup>, using the *bwa* software package<sup>13</sup>. Each read was aligned using the software package, then alignments were paired. Alignment files were then subsetted to consider only proper pairs with a quality score greater than 12<sup>14</sup>.

**Variant identification.** Variable sites were generated, regardless of coverage depth assuming a haploid genome using the *bcftools* software package<sup>14</sup>. Variable sites were then filtered for quality and coverage. Finally, the consensus genome was generated (CoSa suite, Pacific Biosciences, Menlo Park, CA). Consensus genomes assessed with *pangolin* and *nextclade*<sup>15</sup>.

**Ad-hoc analysis.** Custom scripts were used to calculate sequencing, effort, and the percentage of reads used to assemble de novo genomes and base pair coverage. These can be found at (<https://github.com/curative/bioinformatics>).

**SARS-CoV-2 prevalence in Los Angeles and Riverside counties.** Pooled de-identified SARS-CoV-2 test from Los Angeles County and Riverside County was plotted over time to assess SARS-CoV-2 prevalence during our study period and compared to country data that included test results from all testing sites that partnered with the county. Results are displayed in a figure of percentage positivity, i.e., the number of positive tests over a total number of tests conducted.

**Analysis of consensus sequences mutations and identifying similar isolates.** A lineage comparison was done using the *Outbreak.info* resource (Scripps Research, La Jolla, CA 92037). A search for genomic sequences similar to the identified SARS-CoV-2 isolates was performed using Nucleotide BLAST 2.6.0+. Iso-

lated were compared to all sequences available on GISAID (<https://www.epicov.org/epi3/cfrontend#2c08bd>). Data underwent alignment to identify gaps in generated consensus sequences and to match them with positions in amino acid sequences carrying hallmark mutations using Geneious Prime software (Geneious, Auckland, New Zealand).

**Data availability and materials.** Results were submitted to the Global Influenza Surveillance and Response System (GISAID) EpiCoV database for widespread data sharing and surveillance, a public surveillance service<sup>16,17</sup>. All results were submitted to the GISAID EpiCoV database.

The accession numbers that were added are listed in Supplemental Table 5 and are publicly available. The datasets generated and/or analyzed during the current study are available in the GISAID EpiCoV database repository, <https://gisaid.org/>.

**Human subjects.** The study was approved by Advarra Institutional Review Board under Pro00053729 on May 10, 2021. All research was performed in accordance with relevant regulations in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants. No vulnerable groups were included in this study.

## Results

From May 27, 2021 to September 9, 2021, 503 SARS-CoV-2 PCR positive participants were enrolled and underwent specimen collection. Participants were enrolled over several months of the study period (Supplemental Table 1). Over the study period, laboratory percentage positivity data was compared to county positivity, the positivity of SARS-CoV-2 tests were similar temporally (Fig. 1b). Of the 503 participants, 238 (47.3%) participants were women, 329 (63.6%) participants were vaccinated, and 221 (43.9%) participants were of Hispanic or Spanish origin (Table 1). Of the cohort, 496 (98.6%) participants had symptoms at time of collection with congestion, cough, and fatigue being most common. Among the 503 samples, 443 (88.1%) nasal specimens and 353 (70.2%) oral specimens yielded positive sequencing results.

Over our study period, the prevalence of the Alpha variant of SARS-CoV-2 decreased (initially 23.1% [95% confidence interval (95% CI) 0–0.49%] to 0% [95% CI 0.0–0.0%]) as the prevalence of the Delta variant of SARS-CoV-2 increased (initially 33.3% [95% CI 0.0–100.0%] to 100.0% [95% CI 100.0–100.0%]; Fig. 1a,c, and Supplemental Table 2). SARS-CoV-2 lineages AY.4 and AY.12 made up most of the identified Delta variant. Changes of SARS-CoV-2 lineage were similar among those vaccinated and unvaccinated (Supplemental Fig. 1a,b). Cycle threshold values were similar between those vaccinated and unvaccinated at time of collection (Supplemental Table 4).

As of 29 October 2021, there were 2,110,018 reported sequences for all SARS-CoV-2 Delta lineages, with 829,064 reported sequences belonging to the B.1.617.2 lineage. Over 63% of samples originally reported as B.1.617.2 had at least one gap in their consensus sequence. The most common mutation was in the G29422T position.

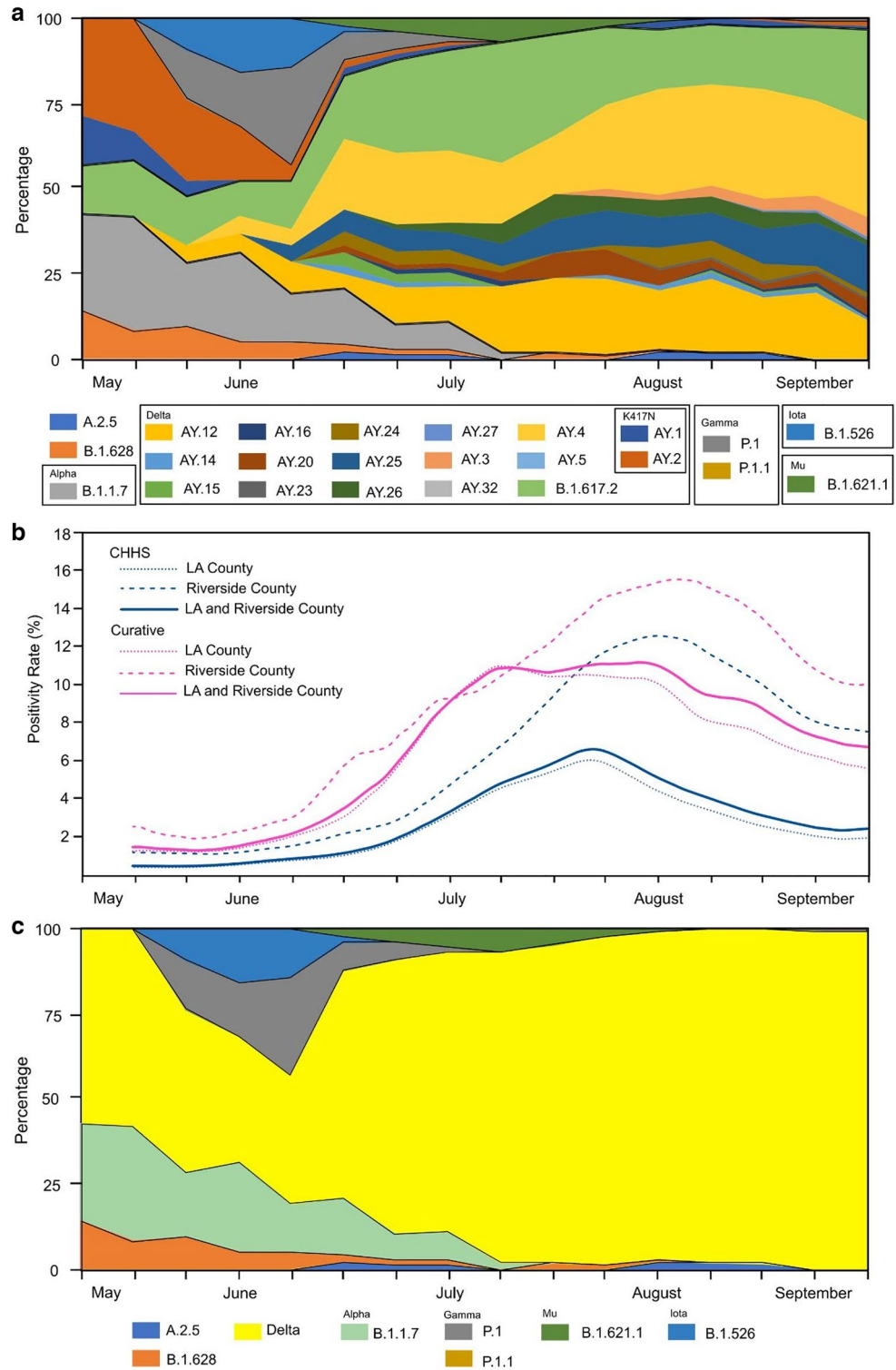
During our period of surveillance, we also found a unique case of a non-identifiable isolate (hCoV-19/USA/CA-Curative-707962712299/2021) collected on July 7, 2021 sequenced on July 14, 2021 (Supplemental Table 2, Supplemental Figs. 2, 3a,b, and 4). The non-identifiable isolate had 95.79% non-N coverage. The non-identifiable isolate carried hallmark S protein mutations from Alpha, Beta, Delta, and Mu lineages of SARS-CoV-2 simultaneously, as well as rare mutations detected in AY.20 sub-lineage of Delta (Supplemental Figs. 2, 3a,b, and 4). Initially the non-identifiable isolate was classified as the B.1.621.1 lineage (Mu) using PANGO version 1.2.36. However, there is no known lineage that matches the non-identifiable isolate with PANGO version 3.1.11. BLAST alignment of the isolate consensus against the GISAID SARS-CoV-2 sequences revealed that this isolate is highly similar to >20 cases detected in Mexico starting June 13, 2021 and 7 isolates sequenced in the United States either by this private laboratory or the United States Centers for Disease Control and Prevention (Supplemental Fig. 3a,b).

## Discussion

We found that outpatient SARS-CoV-2 variant surveillance could be conducted in a private laboratory in a timely and accurate manner. SARS-CoV-2 positivity rates from specimens tested in a private laboratory were similar to county level data collected during the same periods. Among a sample of people that tested positive for SARS-CoV-2, we observed that the Alpha variant of SARS-CoV-2 was most prevalent in Los Angeles and Riverside County in May 2021, however the Delta variant of SARS-CoV-2, mainly the AY.4 and AY.12 strains, became dominant over a period of weeks. Among the isolates identified as a SARS-CoV-2 Delta variant, a large number of isolates carried mutations classifying them as a sub-lineages of Delta, which is evidence of continued mutation. When compared to publicly available data on SARS-CoV-2 variants, our surveillance sampling found similar proportions of variants of concern in our outpatient samples over time<sup>18</sup>.

The identification of new SARS-CoV-2 variants in a timely manner is critical to public health. While it is hard to prognosticate the future, it is possible to establish a method to prioritize research of new identified virus lineages that carry novel mutations on genetic coding segments of key proteins, like the SARS-CoV-2 spike protein<sup>19,20</sup>. Faster identification of new SARS-CoV-2 variants of concern and understanding the rates in their change of prevalence could be critical predictors of new waves of SARS-CoV-2 and met with changes in public health recommendations. This study demonstrates that private laboratories can have a role in the surveillance of SARS-CoV-2 variants of concern.

The sheer number of people who have been infected and the total SARS-CoV-2 infected person-time has led to the rapid evolution of SARS-CoV-2. Local epidemics of populous areas creates a situation in which many new



**Figure 1.** (a) SARS-CoV-2 expanded lineage data over time from a cohort of outpatient participants in Los Angeles County and Riverside County, California, May 27, 2021 to September 9, 2021 (n = 503). (b) Total SARS-CoV-2 positivity rates detected in a private laboratory in Los Angeles County and Riverside County, California, May 27, 2021 to September 9, 2021 (n = 503). (c) SARS-CoV-2 lineage data by variant over time from a cohort of outpatient participants in Los Angeles County and Riverside County, California, May 27, 2021 to September 9, 2021 (n = 503). \*Lineages with the K417N substitution were replaced by different Delta lineages. Data was smoothed over three consecutive weeks.

Variable	Mean/number	Standard deviation or percent
Age, years	40.9	12.6
Gender, female	238	47.3%
Days to collection	2.8	0.7
<b>Ethnicity</b>		
Not disclosed	25	5.0%
Black or African American	1	0.2%
Hispanic or Spanish origin	221	43.9%
Not Hispanic or Spanish origin	208	41.4%
Prefer not to share	48	9.5%
<b>Vaccinated</b>		
No	182	36.2%
Yes	320	63.6%
Declined to disclose	1	0.2%
<b>Vaccine manufacturer</b>		
J&J	56	16.6%
Moderna	98	29.1%
Pfizer	179	53.1%
Pfizer and AstraZeneca	1	0.3%
Sinopharm	3	0.9%
<b>Symptoms</b>		
Congestion or runny nose	375	75.6%
Cough	368	74.2%
Fatigue	343	69.2%
Headache	301	60.7%
Muscle or body aches	290	58.5%
New loss of taste or smell	289	58.3%
Fever or chills	220	44.4%
Sore throat	216	43.5%
Shortness of breath	125	25.2%
Diarrhea	105	21.2%
Nausea or vomiting	80	16.1%
Bleeding	1	0.2%
Eye redness or dry mouth	1	0.2%
Stomach pain	1	0.2%

**Table 1.** Demographic characteristics, clinical symptoms and SARS-CoV-2 lineage data from a cohort of outpatient participants in Los Angeles County and Riverside County, California, May 27, 2021 to September 9, 2021 (n = 503).

mutations can form due to the large amount of viral replication over a short period of time. It is essential that all global SARS-CoV-2 epidemics are controlled to limit the rate of new SARS-CoV-2 mutations.

Temporally, it can be observed that the Delta variant of SARS-CoV-2 arose following *en masse* vaccination efforts. There are many reported cases of breakthrough SARS-CoV-2 infections among people who are fully vaccinated. Many of those people had high viral loads (cycle threshold values less than 30), and with cycle threshold values similar when comparing those who were vaccinated to those who were not vaccinated<sup>21,22</sup>. Reassuringly, most breakthrough infections were mild or asymptomatic<sup>23</sup>. Additionally, only small differences were observed in vaccine effectiveness against symptomatic disease or death when comparing the Delta to Alpha variant with the BNT162b2 vaccine (93.7% with Alpha and 88.0% for Delta) and ChAdOx1 nCoV-19 vaccine (74.5% with Alpha and 67.0% for Delta)<sup>24</sup>. Due to the erosion of vaccine efficacy against the SARS-CoV-2 Delta variant, it is possible that vaccinations may have played a role in the selective pressure for the Delta variant prevalence in areas with high vaccination rates<sup>7,9,20,25</sup>.

In our efforts to monitor SARS-CoV-2 variants of concern, we sequenced a non-identifiable isolate of SARS-CoV-2, the hCoV-19/USA/CA-Curative-707962712299/2021 isolate. The non-identifiable isolate simultaneously carried classic S protein mutations present in the Delta variant, while also displaying hallmark S protein mutations observed in Alpha, Beta, Mu and Kappa variants of SARS-CoV-2. The non-identifiable isolate carried seven S protein mutations prevalent in Mu variant, notably the S:R346K amino acid substitution in the Receptor Binding Domain and two mutations of concern/interest, S:E484K and S:N501Y, in the Receptor-Binding Motif<sup>26</sup>, which are also normally absent in Delta and Delta plus variants. Also, the non-identifiable isolate carried 7 mutations prevalent in AY.20 sub-lineage of Delta, including P681R mutation special for Delta lineage and known to

facilitate the spike protein cleavage, enhance cell-level infectivity and pathogenicity<sup>19</sup>. It carried other common Delta mutations (S:T19R, S:T478K) that are not prevalent in Mu variants. Notably, the isolate carries mutations affecting epitopes for all three main classes of neutralizing antibodies (Class 1-N501Y, T478K; Class 2-E484K, L452; and Class 3-R346K), which brings concerns that this isolate might have evolutionary advantages similar to the Delta variant of SARS-CoV-2 with its ability to evade the immune system of vaccinated persons and have increased infectivity<sup>7,27</sup>.

The study had the following limitations. Given that our study population consisted of people undergoing outpatient SARS-CoV-2 testing, it is possible that the variants we identified are less likely to cause critical illness due to our sampling methodology. The study was conducted in Los Angeles and Riverside Counties over a short period of time, therefore there are temporal and geographic constraints of this study. Reclassification of SARS-CoV-2 variants may change our results. We believe it is likely that the isolates we sequenced will be further reclassified as new lineages or have features of interest/concern as governing bodies determine new SARS-CoV-2 variants of concern in the future.

## Conclusion

We found that outpatient SARS-CoV-2 variant surveillance could be conducted in a private laboratory in a timely and accurate manner. A higher proportion of nasal specimens yielded results than oral specimens. We observed that among a sample of people that tested positive for SARS-CoV-2, initially the Alpha variant was most prevalent in May, however the Delta variant of SARS-CoV-2 quickly became the most prevalent over a period of weeks. When compared to publicly available data on SARS-CoV-2 variants, our surveillance sampling found similar proportions of variants of concern in our outpatient samples over time. We identified a potentially new non-identifiable isolate was identified in our study cohort through surveillance measures. This study demonstrates that timely outpatient SARS-CoV-2 variant surveillance could be used for public health efforts to identify changes in SARS-CoV-2 strains in local epidemics. Government agencies should engage private laboratories in the surveillance of diseases that threaten the public's health to supplement national disease reporting networks.

Received: 17 November 2021; Accepted: 27 September 2022

Published online: 11 October 2022

## References

1. Baric, R. S. Emergence of a highly fit SARS-CoV-2 variant. *N. Engl. J. Med.* **383**, 2684–2686. <https://doi.org/10.1056/NEJMcibr2032888> (2020).
2. Luring, A. S. & Malani, P. N. Variants of SARS-CoV-2. *JAMA* <https://doi.org/10.1001/jama.2021.14181> (2021).
3. Tao, K. *et al.* The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat. Rev. Genet.* <https://doi.org/10.1038/s41576-021-00408-x> (2021).
4. Abdool Karim, S. S. & de Oliveira, T. New SARS-CoV-2 variants—Clinical, public health, and vaccine implications. *N. Engl. J. Med.* **384**, 1866–1868. <https://doi.org/10.1056/NEJMc2100362> (2021).
5. Del Rio, C., Malani, P. N. & Omer, S. B. Confronting the Delta variant of SARS-CoV-2, Summer 2021. *JAMA* <https://doi.org/10.1001/jama.2021.14811> (2021).
6. Wink, P. L. *et al.* First identification of SARS-CoV-2 Lambda (C.37) variant in Southern Brazil. *Infect. Control Hosp. Epidemiol.* <https://doi.org/10.1017/ice.2021.390> (2021).
7. Planas, D. *et al.* Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **596**, 276–280. <https://doi.org/10.1038/s41586-021-03777-9> (2021).
8. Walensky, R. P., Walke, H. T. & Fauci, A. S. SARS-CoV-2 variants of concern in the United States—Challenges and opportunities. *JAMA* **325**, 1037–1038. <https://doi.org/10.1001/jama.2021.2294> (2021).
9. Mlcochova, P. *et al.* SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* <https://doi.org/10.1038/s41586-021-03944-y> (2021).
10. World Health Organization. *Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health.* (World Health Organization, 2021).
11. Zhang, X. *et al.* pTrimmer: An efficient tool to trim primers of multiplex deep sequencing data. *BMC Bioinform.* **20**, 236. <https://doi.org/10.1186/s12859-019-2854-x> (2019).
12. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269. <https://doi.org/10.1038/s41586-020-2008-3> (2020).
13. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595. <https://doi.org/10.1093/bioinformatics/btp698> (2010).
14. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience.* <https://doi.org/10.1093/gigascience/giab008> (2021).
15. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407. <https://doi.org/10.1038/s41564-020-0770-5> (2020).
16. Maxmen, A. One million coronavirus sequences: Popular genome site hits mega milestone. *Nature* **593**, 21. <https://doi.org/10.1038/d41586-021-01069-w> (2021).
17. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data—From vision to reality. *Euro Surveill.* <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494> (2017).
18. Wadman, M. U.S. ups coronavirus surveillance. *Science* **373**, 1184. <https://doi.org/10.1126/science.acx9045> (2021).
19. Callaway, E. The mutation that helps Delta spread like wildfire. *Nature* **596**, 472–473. <https://doi.org/10.1038/d41586-021-02275-2> (2021).
20. Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **19**, 409–424. <https://doi.org/10.1038/s41579-021-00573-0> (2021).
21. Brown, C. M. *et al.* Outbreak of SARS-CoV-2 infections, including COVID-19 vaccine breakthrough infections, associated with large public gatherings—Barnstable County, Massachusetts, July 2021. *MMWR Morb. Mortal. Wkly. Rep.* **70**, 1059–1062. <https://doi.org/10.15585/mmwr.mm7031e2> (2021).
22. Hetemaki, I. *et al.* An outbreak caused by the SARS-CoV-2 Delta variant (B.1.617.2) in a secondary care hospital in Finland, May 2021. *Euro Surveill.* <https://doi.org/10.2807/1560-7917.ES.2021.26.30.2100636> (2021).
23. Bergwerk, M. *et al.* Covid-19 breakthrough infections in vaccinated health care workers. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa2109072> (2021).

24. Lopez Bernal, J. *et al.* Effectiveness of Covid-19 vaccines against the B.1.617.2 (Delta) variant. *N. Engl. J. Med.* **385**, 585–594. <https://doi.org/10.1056/NEJMoa2108891> (2021).
25. Urie, K. *et al.* Neutralization of the SARS-CoV-2 Mu variant by convalescent and vaccine serum. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMc2114706> (2021).
26. Laiton-Donatoa, K. *et al.* Characterization of the emerging B.1.621 variant of interest of SARS-CoV-2. *medRxiv* (2021).
27. Shrestha, L. B., Tedla, N. & Bull, R. A. Broadly-neutralizing antibodies against emerging SARS-CoV-2 variants. *Front. Immunol.* **12**, 752003. <https://doi.org/10.3389/fimmu.2021.752003> (2021).

## Acknowledgements

Elizabeth Scheutzle, Aaliyah Bussanich, Maysoun Sabbagh, Teresa Arellano, and Nina Nirema made contributions to this project.

## Author contributions

N.K. wrote the primary manuscript and performed statistics and analysis. E.K., L.L., M.H., J.C.I., G.W., Z.X., V.S., and J.D.K. edited the draft. E.K., M.H., and J.C.I. performed the variant identification and figure development. J.C., M.G., S.Q., N.S., E.S., G.W., Z.X., and V.S. performed sequencing and developed variant identification techniques. N.K., L.L., M.B., M.G., V.H., and M.S. performed the clinical study.

## Funding

Curative Inc. and the Departments of Medicine and Population and Public Health Sciences and the COVID-19 Pandemic Research Center, University Southern California, Keck School of Medicine, University Southern California, Los Angeles, CA, USA.

## Competing interests

N.K. is a consultant for Curative. E.K., L.K., M.H., J.C.I.E., M.B., J.C., M.G., S.Q., N.S., E.S., M.S., G.W., Z.X. and V.S. are employed by Curative. J.D.K. is an independent consultant and serves as the Medical Director of Curative.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-21481-0>.

**Correspondence** and requests for materials should be addressed to N.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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) 2022