# Local monitoring of SARS-CoV-2 Variants in Two Large California Counties in 2021

**Noah Kojima** ( nkojima@ucla.edu )
University of California Los Angeles https://orcid.org/0000-0002-3667-9719
**Eugenia Khorosheva**
Curative
**Lauren Lopez**
Curative
**Mikhail Hanewich-Hollatz**
Curative
**J. Cesar Ignacio-Espinoza**
Curative
**Matthew Brobeck**
Curative
**Janet Chen**
Curative
**Matthew Geluz**
Curative
**Victoria Hess**
Curative
**Sophia Quasem**
Curative
**Nabjot Sandhu**
Curative
**Elias Salfati**
Curative
**Maria Shacreaw**
Curative
**George Way**
Curative
**Zhiyi Xie**
Curative
**Vladimir Slepnev**
Curative
**Jeffrey Klaunser**

University of Southern California

**Article**

**Keywords:** COVID-19, variant surveillance, strain epidemiology

**Posted Date:** November 17th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-1061555/v1

# Abstract

**Introduction:** 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. The initial government response to COVID-19 failed to engage the private sector. Engaging the private sector could have bolstered the national capacity to process diagnostic tests and track variants of concern for disease for public health surveillance. 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.

**Methods**: Study enrollment was offered to adults aged 18 years or older in Los Angeles County and Riverside County in California who recently tested positive for SARS-CoV-2 by polymerase chain reaction (PCR) with a cycle threshold value less than or equal to 30 cycles. Trained healthcare workers directly observed self-collection of oral fluid or anterior nares specimens within 5 days of study enrollment. Specimens were transported and stored at 8°C or cooler. RNA was extracted from samples. Samples underwent library preparation and were sequenced. Sequencing data were filtered by quality criteria. High-quality genomic data were analyzed to identify SARS-CoV-2 lineages. Participant and genomic data were analyzed using statistical tools and visualized with toolkits. The study was approved by Advarra Institutional Review Board (Pro00053729).

**Results:** From May 27, 2021 to September 9, 2021, 503 participants were enrolled and underwent specimen collection. Of those enrolled, there were 238 (47.3%) females, 329 (63.6%) vaccinated, and 221 (43.9%) of Hispanic or Spanish origin. Of the cohort, 496 (98.6%) had symptoms at the time of collection. Among the 503 participants, 443 (88.1%) nasal specimens and 353 (70.2%) oral specimens yielded 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% to 0.49%] to 0% [95% CI: 0.0% to 0.0%]) as the prevalence of the Delta variant of SARS-CoV-2 increased (initially 33.3% [95% CI: 0.0% to 100.0%] to 100.0% [95% CI: 100.0% to 100.0%]). A strain that carried mutations of both Delta and Mu was identified.

**Conclusion:** We found that outpatient SARS-CoV-2 variant surveillance could be conducted in private laboratory in a timely and accurate manner. The prevalence of different variants changed over time. A higher proportion of nasal specimens yielded results when compared to oral specimens. Timely outpatient SARS-CoV-2 variant surveillance could be used for public health efforts to identify changes in SARS-CoV-2 strain epidemiology in local areas. Government agencies should engage private laboratories in the surveillance of diseases that threaten the public's health to supplement national disease reporting networks.


# Introduction

Coronavirus disease 2019 (COVID-19) cases, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), continue. Like many viruses, SARS-CoV-2 mutates.[1] When those mutations

occur in important parts of key proteins that increase chance of spread, decrease susceptibility to treatments, or effect vaccine efficacy, those are classified as variant of concern.[2,3]

Following the initial reports of SARS-CoV-2 infections in China and Italy, the United States failed to prepare for a pandemic—the SARS-CoV-2 assay developed by the US Centers for Disease Control and Protection was flawed, [4] the supply-chain for diagnostic and medical supplies was not shored,[5] and the public health system was neither prepared for the volume of positive SARS-CoV-2 tests nor surveillance for new variants of SARS-CoV-2.[6] To have effectively combated SARS-CoV-2, public-private partnerships were not only needed for the development of effective vaccines,[7] but also in the rapid expansion of diagnostic testing and variant surveillance early in the pandemic period.[8]

Following the initial wave of the wildtype strain of SARS-CoV-2, the first identified variant of concern was the Alpha variant (UK variant), which spreads approximately 50% faster than the wildtype strain.[2] Following the Alpha variant of concern, the Beta (South Africa) and Gamma (Brazilian) variants of concern were identified.[9] Currently, the Delta variant is the most prevalent SARS-CoV-2 variant of concern in the United States,[10] with new variants actively being identified.[11] Compared to the wildtype strain of SARS-CoV-2, the Delta variant is more infectious and might escape immune responses from vaccination.[2,12–14]

Due to the major public health concern of SARS-CoV-2 and identified variants, surveillance programs are needed to monitor SARS-CoV-2 variants to inform public health efforts. With the development of new genomic sequencing tools, it is possible for genomic data to be used to inform those public health responses. 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.[15] 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 recently tested positive for SARS-CoV-2 by reverse transcription polymerase chain reaction (RT-PCR) 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 sample collection. Subjects considered vulnerable including 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 eligible participant prior to enrollment in the study and any specimen collection.

Individuals meeting eligibility criteria and providing written consent were enrolled in the study. Subject symptom status and vaccination information was gathered from participants at enrollment.

# Specimen Collection and Transport

Trained healthcare workers visited subject homes and instructed participants to self-collect oral fluid and anterior nares specimens under direct observation. 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°C to 8°C within 5 hours 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 chaotropic agents/silica-based methods.

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 KingFisher™ Flex, Thermo Scientific, Waltham, MA) were used. RNA was eluted in 60 mL of 10 mM Tris (pH 7.4).

# Library preparation and sequencing

The libraries were prepared using Illumina COVIDSeq protocol (Illumina, Inc., San Diego, CA). On a single 96-well plate, 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. Next, 96 indexed sample libraries from each plate were pooled together and quantified using a fluorometer (Qubit 3.0, Invitrogen, Waltham, MA). Four 96-well plates were combined at equimolar concentrations to a total of 384 samples and sequenced. Dilution and loading were performed as per the manufacturer's instructions (Illumina, Inc., San Diego, CA). Dual-indexed paired-end sequencing was performed for 100 cycles or 200 cycles to get a deeper sequencing depth. Sequencing aimed to have 1 to 2 million reads per sample (NextSeq2000, Illumina, Inc., San Diego, CA).

# Quality control of reads

Paired-end reads were filtered and trimmed to reduce low quality base calls in the analysis and to eliminate the presence of primers and adapters. A minimum quality score of 30 was selected (*pTrimer-1.3.4 -a Primers.bed -q 30 -t pair*).[16]

# Mapping

Paired-end reads were then mapped against the 'Wuhan seafood market pneumonia virus isolate" Wuhan-Hu-1 genome (*Accession number: NC_045512.2*),[17] using *bwa*.[18] Each read was aligned (*bwa aln -t 8 NC_045512.2.fasta*), then alignments were paired with the *sampe* option. Alignment files were then subsetted using samtools view to consider only proper pairs with a quality score larger than 12 (*samtools view -bS -q 12 -f 0X2*).[19]

# Variant calling and consensus genome generation, mutation identification, and variant identification

Variable sites were generated, regardless of coverage depth assuming a haploid genome using bcftools (*bcftools call -mv -Ov*).[19] Variable sites were then filtered for quality (20) coverage (20x) and minimum allele frequency (0.25). Finally, the consensus genome was generated using the VCFcons.py script (*VCFCons.py –input_depth TEMP.depth –input_vcf sample.vcf –vcf_type bcftools -c 10 -f 0.25 -q 20*), part of the CoSa suite (Pacific Biosciences, Menlo Park, CA). Consensus genomes were then run on the command line version of both: pangolin and nextclade.[20]

# 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).

# Generation of SARS-CoV-2 Prevalence in Los Angeles and Riverside Counties

Pooled de-identified SARS-CoV-2 test data collected in Los Angeles County and Riverside County that was collected and resulted in a private laboratory (Curative, San Dimas, CA) was plotted over time.

# Analysis of consensus sequences mutations and identifying similar isolates

A lineage comparison was done using Outbreak.info resource created by Scripps Research. A search for genomic sequences similar to the identified SARS-CoV-2 isolates was performed using Nucleotide BLAST 2.6.0+. Isolated 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).

# Human subjects

The study was approved by Advarra Institutional Review Board under Pro00053729 on May 10, 2021.

## Results

From May 27, 2021 to September 9, 2021, 503 SARS-CoV-2 PCR positive participants were enrolled and underwent specimen collection. When laboratory positivity data was compared to county positivity, the positivity of SARS-CoV-2 tests were similar temporally (Figure 1b). Of those enrolled, there were 238 (47.3%) females, 329 (63.6%) vaccinated, and 221 (43.9%) 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. Of the 503 participants, 443 (88.1%) nasal specimens and 353 (70.2%) oral specimens yielded sequencing results. Participants were entolled over several months of the study period (Supplemental Table 1).

Over our study period, the prevalence of the Alpha variant of SARS-CoV-2 decreased (initially 23.1% [95% confidence interval (95% CI): 0% to 0.49%] to 0% [95% CI: 0.0% to 0.0%]) as the prevalence of the Delta variant of SARS-CoV-2 increased (initially 33.3% [95% CI: 0.0% to 100.0%] to 100.0% [95% CI: 100.0% to 100.0%]; Figure 1a, 1c, 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 Figure 1a and 1b). 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 3, Supplemental Figure 2, 3ab, 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 Figure 2, 3ab, 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 Stated either by this private laboratory or the United States Centers for Disease Control and Prevention (Supplemental Figure 3a and 3b).

*Reporting*

Results were submitted to Global Influenza Surveillance & Response System (GISAID) EpiCoV database for widespread data sharing and surveillance, a public surveillance service.[21,22] The accession numbers were added to GISAID (Supplemental Table 5).

# 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 initially 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.[23]

After chronic neglect of defunding, the public health infrastructure of the United States was ill equipped for the COVID-19 pandemic.[24] Despite earlier warnings of SARS-CoV-2 epidemics in early 2020,[25,26] the United States was not prepared for SARS-CoV-2 testing.[4] Despite the poor state of public health infrastructure of the United States for a pandemic, hundreds of thousands of deaths could have potentially been avoided if public-private partnerships were developed for make up for the dearth in diagnostic testing and health workforce capacity.[27]

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 when new mutations are discovered on genetic coding segments of key proteins, like the SARS-CoV-2 spike protein.[28,29] 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 have a role in the surveillance of SARS-CoV-2 variants of concern.

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.[30,31] Reassuringly, most breakthrough infections were mild or asymptomatic.[32] 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).[33] Due to the erosion gf 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.[12,14,29,34]

In our efforts to monitor SARS-CoV-2 variants of concern, we sequenced a non-identifiable isolate of SARS-CoV-2, hCoV-19/USA/CA-Curative-707962712299/2021. 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,[35] 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.[28] 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.[12,36]

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 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.

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 if those patients required hospitalization. The study was conducted in Los Angeles and Riverside Counties therefore temporal changes may differ from other outbreaks across the country and world. Reclassification of SARS-CoV-2 variants may change our results: When identifying the Delta strain of SARS-CoV-2, some of the Delta lineages might be further mutated sub-lineages of Delta that carry hallmark mutations of the Delta variant not covered well in generated consensuses. When re-running our samples though the latest version of PangoLearn, sub-lineages of the Delta variant were observed in our study sample since 2020. These samples were not identified as special sub-lineages or variants of concern until variants were reclassified. 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. 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.

# References


1. Baric, R. S. Emergence of a Highly Fit SARS-CoV-2 Variant. *N Engl J Med* **383**, 2684–2686, doi:10.1056/NEJMcibr2032888 (2020).
2. Lauring, A. S. & Malani, P. N. Variants of SARS-CoV-2. *JAMA*, doi: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*, doi:10.1038/s41576-021-00408-x (2021).
4. Schneider, E. C. Failing the Test - The Tragic Data Gap Undermining the U.S. Pandemic Response. *N Engl J Med* **383**, 299–302, doi:10.1056/NEJMp2014836 (2020).
5. Mirchandani, P. Health Care Supply Chains: COVID-19 Challenges and Pressing Actions. *Ann Intern Med* **173**, 300–301, doi:10.7326/M20-1326 (2020).
6. Maxmen, A. Has COVID taught us anything about pandemic preparedness? *Nature* **596**, 332–335, doi:10.1038/d41586-021-02217-y (2021).
7. Bok, K., Sitar, S., Graham, B. S. & Mascola, J. R. Accelerated COVID-19 vaccine development: milestones, lessons, and prospects. *Immunity* **54**, 1636–1651, doi:10.1016/j.immuni.2021.07.017 (2021).
8. Krijger, P. H. L. *et al.* A public-private partnership model for COVID-19 diagnostics. *Nat Biotechnol* **39**, 1182–1184, doi:10.1038/s41587-021-01080-6 (2021).
9. 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, doi:10.1056/NEJMc2100362 (2021).
10. Del Rio, C., Malani, P. N. & Omer, S. B. Confronting the Delta Variant of SARS-CoV-2, Summer 2021. *JAMA*, doi:10.1001/jama.2021.14811 (2021).
11. Wink, P. L. *et al.* First identification of SARS-CoV-2 Lambda (C.37) variant in Southern Brazil. *Infect Control Hosp Epidemiol*, 1–7, doi:10.1017/ice.2021.390 (2021).
12. Planas, D. *et al.* Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **596**, 276–280, doi:10.1038/s41586-021-03777-9 (2021).
13. 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, doi:10.1001/jama.2021.2294 (2021).
14. Mlcochova, P. *et al.* SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature*, doi:10.1038/s41586-021-03944-y (2021).

15. Organization, W. H. (World Health Organization, Geneva, 2021).
16. Zhang, X. *et al.* pTrimmer: An efficient tool to trim primers of multiplex deep sequencing data. *BMC Bioinformatics* **20**, 236, doi:10.1186/s12859-019-2854-x (2019).
17. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269, doi:10.1038/s41586-020-2008-3 (2020).
18. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595, doi:10.1093/bioinformatics/btp698 (2010).
19. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, doi:10.1093/gigascience/giab008 (2021).
20. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* **5**, 1403–1407, doi:10.1038/s41564-020-0770-5 (2020).
21. Maxmen, A. One million coronavirus sequences: popular genome site hits mega milestone. *Nature* **593**, 21, doi:10.1038/d41586-021-01069-w (2021).
22. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* **22**, doi:10.2807/1560-7917.ES.2017.22.13.30494 (2017).
23. Wadman, M. U.S. ups coronavirus surveillance. *Science* **373**, 1184, doi:10.1126/science.acx9045 (2021).
24. Alfonso, Y. N., Leider, J. P., Resnick, B., McCullough, J. M. & Bishai, D. US Public Health Neglected: Flat Or Declining Spending Left States Ill Equipped To Respond To COVID-19. *Health Aff (Millwood)* **40**, 664–671, doi:10.1377/hlthaff.2020.01084 (2021).
25. Guan, W. J. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* **382**, 1708–1720, doi:10.1056/NEJMoa2002032 (2020).
26. Megna, R. First month of the epidemic caused by COVID-19 in Italy: current status and real-time outbreak development forecast. *Glob Health Res Policy* **5**, 43, doi:10.1186/s41256-020-00170-3 (2020).
27. Waitzberg, R. *et al.* Early health system responses to the COVID-19 pandemic in Mediterranean countries: A tale of successes and challenges. *Health Policy*, doi:10.1016/j.healthpol.2021.10.007 (2021).
28. Callaway, E. The mutation that helps Delta spread like wildfire. *Nature* **596**, 472–473, doi:10.1038/d41586-021-02275-2 (2021).
29. Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* **19**, 409–424, doi:10.1038/s41579-021-00573-0 (2021).
30. 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, doi:10.15585/mmwr.mm7031e2 (2021).
31. 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* **26**, doi:10.2807/1560-7917.ES.2021.26.30.2100636

(2021).

32. Bergwerk, M. *et al.* Covid-19 Breakthrough Infections in Vaccinated Health Care Workers. *N Engl J Med*, doi:10.1056/NEJMoa2109072 (2021).
33. 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, doi:10.1056/NEJMoa2108891 (2021).
34. Uriu, K. *et al.* Neutralization of the SARS-CoV-2 Mu Variant by Convalescent and Vaccine Serum. *N Engl J Med*, doi:10.1056/NEJMc2114706 (2021).
35. Laiton-Donatoa, K. *et al.* Characterization of the emerging B.1.621 variant of interest of SARS-CoV-2 *medRxiv* (2021).
36. Shrestha, L. B., Tedla, N. & Bull, R. A. Broadly-Neutralizing Antibodies Against Emerging SARS-CoV-2 Variants. *Front Immunol* **12**, 752003, doi:10.3389/fimmu.2021.752003 (2021).

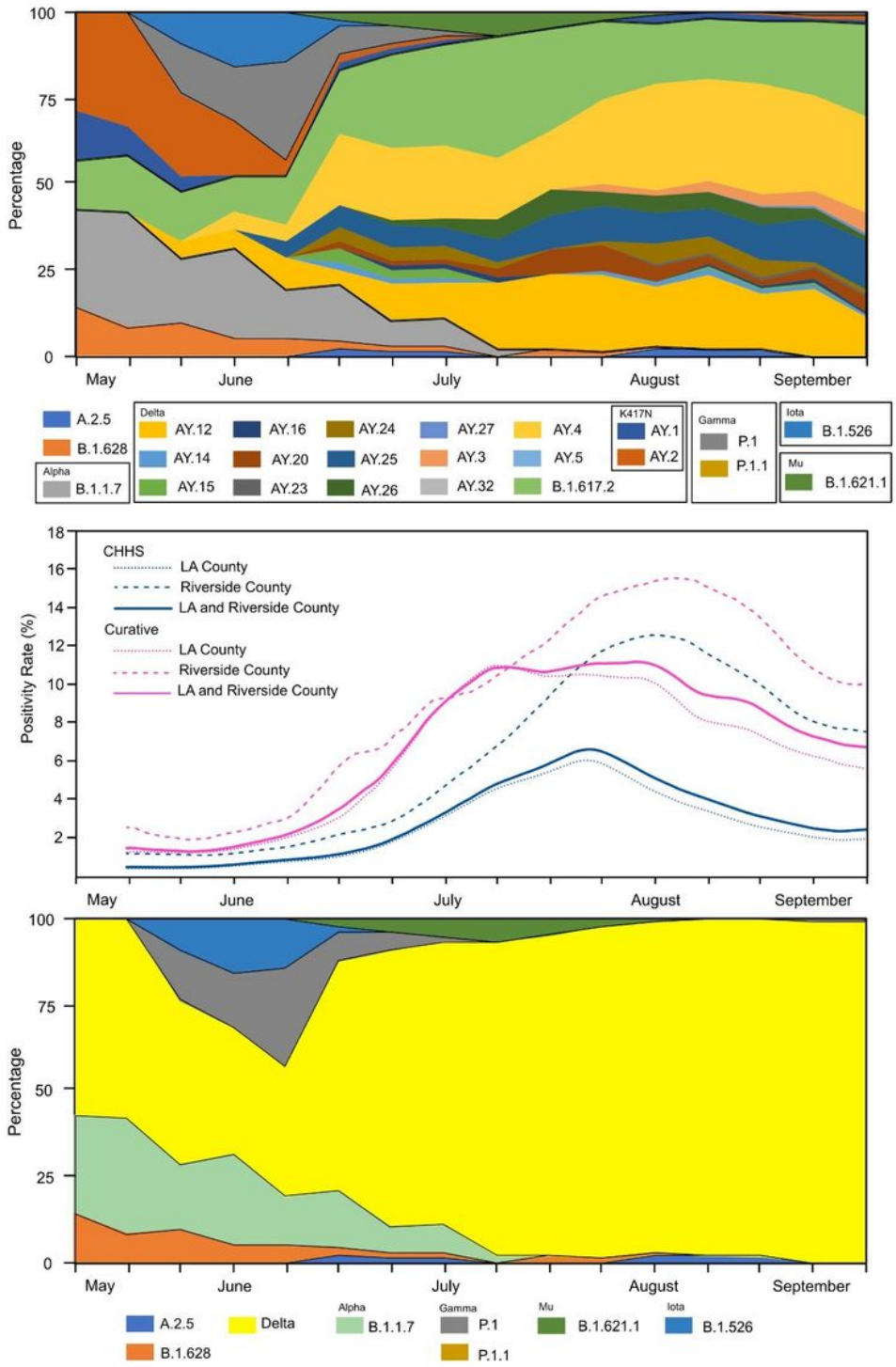
# Tables

**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)

| 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 |
| Ethnicity | | |
| 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% |
| Vaccinated | | |
| No | 182 | 36.2% |
| Yes | 320 | 63.6% |
| Declined to disclose | 1 | 0.2% |
| Vaccine manufacturer | | |
| J&J | 56 | 16.6% |
| Moderna | 98 | 29.1% |
| Pfizer | 179 | 53.1% |
| Pfizer & AstraZeneca | 1 | 0.3% |
| Sinopharm | 3 | 0.9% |
| Symptoms | | |
| 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% |

## Figures




## 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.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalMaterial.docx