Letters

Emergence of a Novel SARS-CoV-2 Variant in Southern California

A spike in coronavirus disease 2019 (COVID-19) has occurred in Southern California since October 2020. Analysis of the severe acute respiratory syndrome coronavirus (SARS-CoV-2) in Southern California prior to October indicated most isolates

➕
Editorial

➕
Supplemental content

originated from clade 20C that likely emerged from New York via Europe early in the pandemic.[1] Since then, novel variants of SARS-CoV-2 including those seen in the UK (20I/501Y.V1/B.1.1.7), South Africa (20H/501Y.V2/B.1.351), and Brazil (P.1/20J/501Y.V3/B.1.1.248) have emerged with the concern of increased infectivity and virulence.[2,3] Thus, we analyzed variants of SARS-CoV-2 in Southern California to establish whether one of these known strains or a novel variant had emerged.

Methods | Regulatory review with waiver of consent was completed by Cedars-Sinai Medical Center (CSMC). From all samples from symptomatic inpatients and ambulatory care (urgent care, primary care, and employee health) that tested positive for SARS-CoV-2 collected from November 22, 2020, to December 28, 2020, at CSMC with cycle threshold values less than 30, a random sample from selected runs and dates within the collection period was sequenced and analyzed (Supplement). In addition, phylogenetic analysis was conducted with CSMC samples and globally representative genomes on January 11, 2021, by utilizing Nextstrain, a collection of open-source tools for visualizing the genetics behind the spread of viral outbreaks.[4] The representative global samples were randomly chosen using a computer algorithm from more than 400 000 available genomes on GISAID (Global Initiative on Sharing All Influenza Data), an open-access global collection of viral genomic data,[5] collected between December 21, 2019, and January 11, 2021 (Supplement).

The proportional prevalence of each clade over time in samples from California as a whole and Southern California specifically and presence of any novel lineages discovered worldwide was calculated using publicly available sequences from GISAID (including samples from CSMC), collected between March 4, 2020, and January 22, 2021. Southern California was defined as including the following counties: Imperial, Kern, Los Angeles, Orange, Riverside, San Bernardino, San Diego, San Luis Obispo, Santa Barbara, and Ventura.

Results | Of 2311 samples at CSMC, 192 were selected and 185 (67 inpatient; 118 outpatient) underwent phylogenetic analysis, along with 1480 representative genomes using Nextstrain. A diverse set of lineages with 2 main clusters was identified (**Figure 1**). The smaller of the 2 clusters was from the 20G lin-

eage and accounted for 22% (40 of 185) of the samples. The larger cluster (36%, 67 of 185) consisted of a novel variant descended from cluster 20C, defined by 5 mutations (ORF1a: I4205V, ORF1b: D1183Y, S: S13I; W152C; L452R) and designated CAL.20C (20C/S:452R; /B.1.429).

Analysis of 10 431 samples from California, including 4829 from Southern California, revealed that CAL.20C was first observed in July 2020 in 1 of 1247 samples from Los Angeles County and not detected in Southern California again until October. Since then, this variant's prevalence has increased in California state and Southern California, where on January 22, 2021, it accounted for 35% (86 of 247) and 44% (37 of 85) of all samples collected in January, respectively (**Figure 2**).

Sequence analysis of 405 871 global samples on GISAID on January 22, 2021, revealed that CAL.20C was only found in Southern California in October 2020 (4 cases). In November 2020, 30 cases were also identified in Northern California and individual cases in 5 additional states. As of January 22, 2021, CAL.20C has been detected in 26 states and other countries (Supplement).

Discussion | A novel variant of SARS-CoV-2, CAL.20C, was identified, which emerged in Southern California contemporaneously with the local surge in cases. Unlike clade 20G, currently the largest reported clade in North America, this strain is defined by 3 mutations in the S-protein characterizing it as a subclade of 20C. The S protein L452R mutation is within a known receptor binding domain that has been found to be resistant to certain spike (S) protein monoclonal antibodies.[6] Because this study was limited to databases of publicly available genomes and a comparatively small set of local samples, the possibility of collection bias cannot be ruled out. Additionally, as clinical outcomes have yet to be established, the functional effect of this strain regarding infectivity and disease severity remains uncertain. Nevertheless, the identification of this novel strain is important to frontline and global surveillance of this evolving virus.

Wenjuan Zhang, PhD
Brian D. Davis, BSc
Stephanie S. Chen, BSc
Jorge M. Sincuir Martinez, BS
Jasmine T. Plummer, PhD
Eric Vail, MD

Author Affiliations: Department of Pathology and Laboratory Medicine, Cedars-Sinai Medical Center, Los Angeles, California (Zhang, Sincuir Martinez, Vail); Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California (Davis, Chen, Plummer).
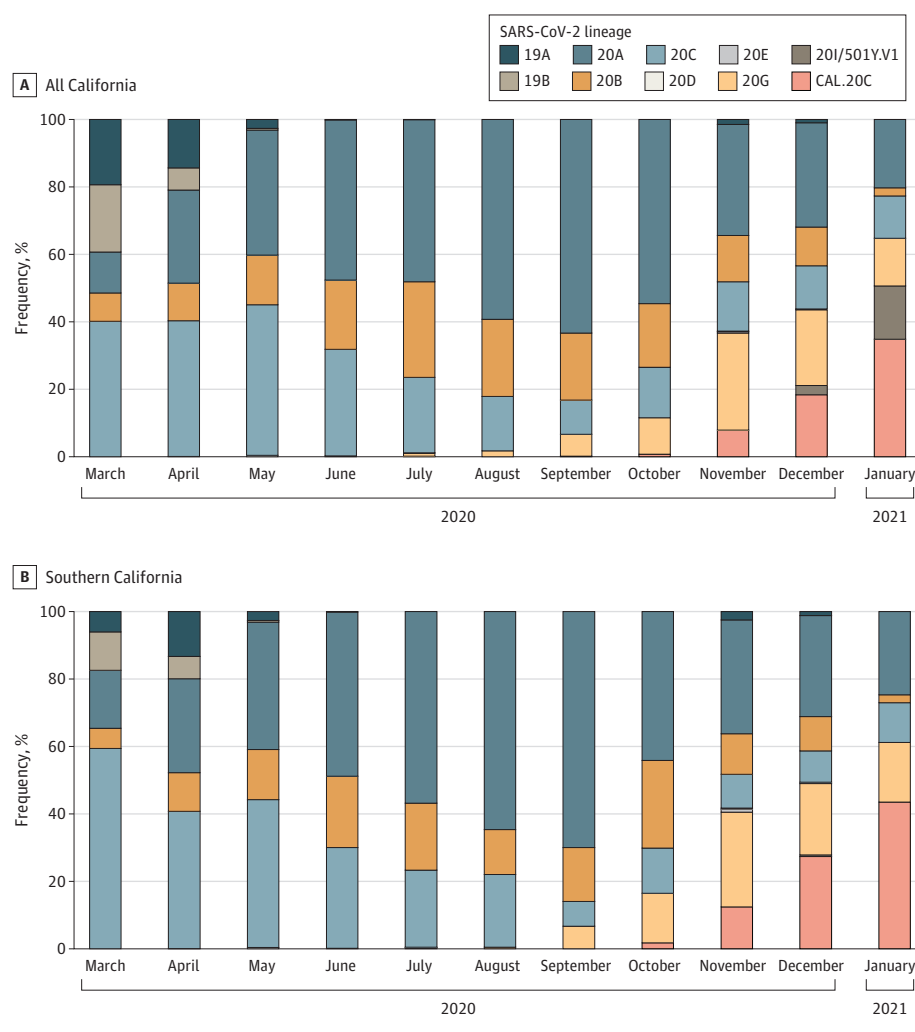
Corresponding Author: Jasmine T. Plummer, PhD, Cedars Sinai Medical Center, 8700 Beverly Blvd, SSB365, Los Angeles, CA 90048 (jasmine.plummer@cshs.org).

Figure 1. Phylogenetic Relationship of CSMC Samples to Global SARS-CoV-2 Genomes
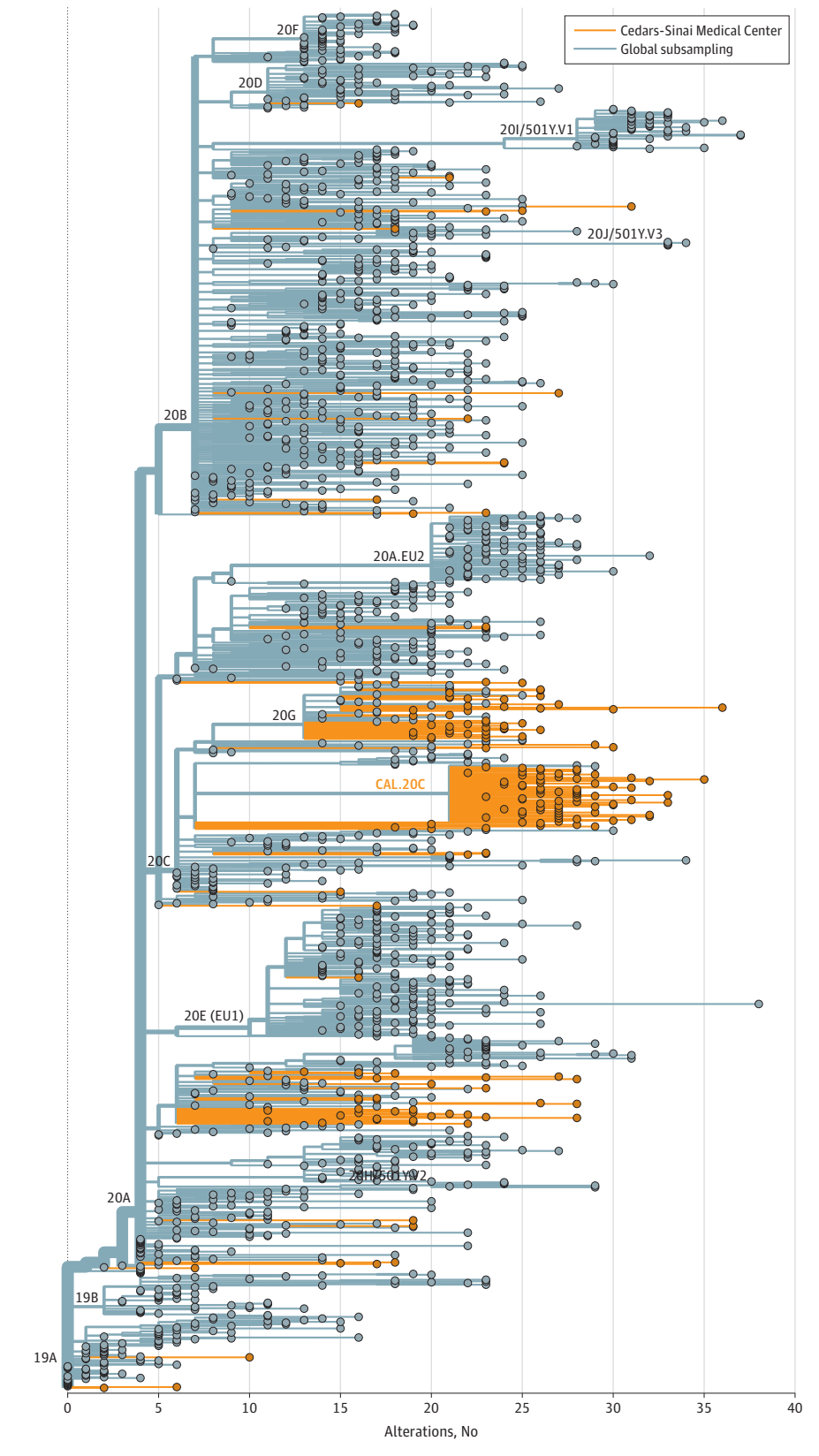
Phylogenetic tree of 185 Cedars-Sinai Medical Center (CSMC) severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) isolates and a global subsampling of 1480 isolates collected from December 2019 to January 2021 reveals a novel subcluster within 20C that share 5 mutations (ORF1a: I4205V, ORF1b: D1183Y, S: S13I; W152C; L452R), designated as CAL.20C (20C/S.452R). The phylogenetic tree shows the relationship of CAL.20C to other circulating lineages. The branch length (x-axis) reflects numbers of mutations accumulated before being discovered, and clades are designated based on Nextstrain nomenclature. The UK variant (501Y.V1), South African variant (501Y.V2), and Brazil variant (501Y.V3) are shown.

Figure 2. Timeline for the Emergence of a Novel Southern California Variant, CAL.20C, Among All SARS-CoV-2 Circulating Variants Observed

Diagrammatic representation of circulating severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variant frequencies. A, Includes 10 431 samples from California state. B, Includes 4829 samples from Southern California.

1. Zhang W, Govindavari JP, Davis BD, et al. Analysis of genomic characteristics and transmission routes of patients with confirmed SARS-CoV-2 in Southern California during the early stage of the US COVID-19 pandemic. *JAMA Netw Open*. 2020;3(10):e2024191. doi:10.1001/jamanetworkopen.2020.24191

2. Lauring AS, Hodcroft EB. Genetic variants of SARS-CoV-2—what do they mean? *JAMA*. Published online January 6, 2021. doi:10.1001/jama.2020.27124

3. Tang JW, Tambyah PA, Hui DS. Emergence of a new SARS-CoV-2 variant in the UK. *J Infect*. Published online December 28, 2020. doi:10.1016/j.jinf.2020.12.024

4. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34(23):4121-4123. doi:10.1093/bioinformatics/bty407

5. Shu Y, McCauley J. GISAID: Global Initiative on Sharing All Influenza Data—from vision to reality. *Euro Surveill*. 2017;22(13):30494. doi:10.2807/1560-7917.ES.2017.22.13.30494

6. Li Q, Wu J, Nie J, et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell*. 2020;182(5):1284-1294.e9. doi:10.1016/j.cell.2020.07.012