Supplemental Online Content

Zhang W, Davis BD, Chen SS, Sincuir Martinez JM, Plummer JT, Vail E. Emergence of a novel SARS-CoV-2 variant in Southern California. *JAMA*. doi:10.1001/jama.2021.1612

eMethods. Diagnostics, Analysis, and Identification of Isolates

This supplemental material has been provided by the authors to give readers additional information about their work.

Supplement

Diagnostics and Sample preparation

Clinical specimens were collected by nasopharyngeal swabs from patients presenting coronavirus disease 2019 (COVID-19) like symptoms. Total nucleic acid was extracted using the QIAamp Viral RNA Mini Kit on the QIASymphony (Qiagen). Following the protocol from our previous study[1], all patients were first assessed by reverse transcriptase real time polymerase chain reaction (RT-qPCR) (Accelerate Technologies, Singapore) for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) viral RNA. Samples were defined as positive with amplification of the targeted region crossing the threshold before 40 cycles. Out of 2311 positive samples collected between 11/22/2020-12/28/2020, 192 samples with cycle threshold (Ct) values below 30 were randomly chosen for sequencing.

Targeted NGS and variant analyses

All samples were processed for NGS and phylogenetic analysis similar to Zhang et al., 2020[1] with the following updates. Using an RNA preparation enrichment tagmentation kit (Illumina), 100 ng of total RNA was processed for first strand and second strand cDNA synthesis. Target enrichment of 200 ng cDNA was performed with the updated viral respiratory panel (RVOP V2) and DNA unique dual indices (Illumina). After enrichment, all samples were pooled, loaded and sequenced on a NovaSeq Illumina platform (75 bp paired-end). All sequencing reads were mapped to SARS-CoV-2 genome (NC_045512.2) and analyzed using our previous established protocol [1]. 185 samples with more than 50% of the genome covered and more than 10× genome depth were retained for phylogenetic analysis. The 185 CSMC samples were assigned with clade names based on Nextstrain[2] naming strategy (01/06/2021) using Nextclade (https://github.com/nextstrain/nextclade). Phylogenetic analysis was conducted combined with 1,480 representative genomes subsampled from global data following methods in previous publication[1] utilizing Nextstrain (https://nextstrain.github.io/ncov/customizing-analysis.html)

GISAID samples identified as CAL.20C

Analysis of data downloaded from GIASID[3] (01/24/2021) revealed that CAL.20C isolates were first identified in Southern California in July 2020. This variant was present in the following locations in November 2020: California, Connecticut, New York, Oregon, Texas, Utah. In December: Arizona, California, Connecticut, Maryland, Michigan, Nevada, New Mexico, New York, Oregon, Utah, Washington, Washington DC, Wyoming, Australia, New Zealand, Israel, Singapore and United Kingdom. In January 2021: Alaska, Arizona, California, Connecticut, Georgia, Hawaii, Maryland, Michigan, New Mexico, Nevada, New York, Rhode Island, South Carolina, Texas, Utah, Washington, Wisconsin, Wyoming, Washington DC,, Australia, Singapore, New Zealand, United Kingdom, Denmark and Israel.

References

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. 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
3. 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