Emerging SARS-CoV-2 Mutations Circulating in the United States

Lauren Brinkac¹, Sheila Diepold², Shane Mitchell¹, Lee F. Kolakowski², William M. Nelson², Katharine Jennings¹
¹ Noblis, Reston, VA; ² Tetracore, Rockville, MD; * contributed equally

ABSTRACT
Coronavirus disease 2019 (COVID-19), the infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a global pandemic of unprecedented scale. Despite significant increases in COVID-19 positive cases in the United States (US), the extent of genomic diversity of SARS-CoV-2 in circulation has yet to be fully elucidated. Utilizing Oxford Nanopore Technologies’ (ONT) MinION™, we can complete whole genome sequences (WGS) of SARS-CoV-2 clinical isolates obtained from confirmed COVID-19 cases throughout the US, were generated and analyzed. Twenty-one isolates were selected to the potential emerging 20G/803A clade and co-harbor distinct mutations such as an additional spike protein mutation, Q913H, located within the helical fusion core of the helical repeat 1 (HR1) region, underscoring the critical need for enhanced genomic surveillance to effectively monitor SARS-CoV-2 evolution.

INTRODUCTION
SARS-CoV-2 emerged in Wuhan, China in late 2019. On March 11, 2020 the World Health Organization (WHO) declared the outbreak a pandemic. The emergence is potentially due to a zoonotic transmission of SARS-CoV-2 to an unknown animal species sold in a wet market in Wuhan. Since the initial whole genome sequence, known as Wuhan-Hu_1, was made public more than 441,773 genomes have been sequenced and deposited in GISAID (https://www.gisaid.org/) as of January 29, 2021. Genomic epidemiology has provided detailed tracking of the rise and spread of SARS-CoV-2 variants around the world. Given the fidelity of RNA viruses, variants are expected. The variants of concern (VOC) result in the species expansion, which mediates entry of the virus into host cells and is a key target for antibody recognition.4 These VOC and their biological implications include increased transmissibility (D614G and N501Y)1,5, altered cell entry (P681H and D677V)6,7, and reduced antibody neutralization (E484Q).8,9 In this study, we analyzed 1,750 of 21 SARS-CoV-2 isolates obtained from COVID-19 positive clinical samples and identified the circulation of isolates in two Midwest US states, a region where isolated strains are not emerging, in order to determine the potential clinical importance of these VOC.

RESULTS
Clinical nephropathylial and nasal samples confirmed to be polymerase chain reaction (PCR) positive for COVID-19 were de-identified and obtained from a Clinical Laboratory Improvement Amendments (CLIA) certified laboratory. The viral RNA was purified from 140 µL of the clinical sample using the QiAamp Viral RNA Mini Kit (Qiagen) following the manufacturer protocol. The CMA was generated using a linear and fluorescent primers mix (New England BioLabs, NEB) and SuperScript IV First Strand Synthesis Kit (Life Technologies). Two multiplex PCR reactions, containing a total of 17 primer pairs, were used to amplify across the SARS-CoV-2 genome.10,11 Each primer pair produces an amplicon approximately 1,900 base pairs (bp) in size with an average of 175 bp overlaps between the amplicons. The PCR was performed using the Q5 High-Fidelity DNA Polymerase (NEB). The amplicons from both primer pools were then pooled, purified, and pooled amplicon was purified using Agencourt AMPure XP beads (Beckman Coulter) and quantified using the Qubit 4.0 Fluorometer and the Qubit double-stranded High Sensitivity kit (Life Technologies). The PCR amplicons were used as templates for a preparative gel extraction (GeneJockey, Bio-Rad) using a custom priming region targeting the N terminal domain of the SARS-CoV-2 S protein, from 29254 to 31284 bp of the genome was amplified targeting the protease (Prr) domain, thus containing the spike (S) protein. A total of 60 ng of PCR amplicon was treated with Ultra II End Prep Enzyme mix (NEB). After end repair, the Native Barcodes 1-24 (ONT) were ligated using the Ultra II Ligation Mix (NEB). Library preparation and sequencing on the ONT MinION using the Sequencing Kit 1D, SQK-LSK 109 (ONT). Up to 24 samples, including a positive control, were pooled for sequencing on the same flow cell. Sequencing was performed on the MinION using the R9.4.1 flow cell for 4 to 12 hours, depending on the number of samples.

Raw nanopore signal was processed using ONT’s Guppy basecaller in high accuracy mode using a single NVIDIA Tesla V100 GPU. The basecalled reads were demultiplexed using ONT’s Guppy barcode to biointformatically separate the reads into their appropriate samples. Unclassifiable reads were discarded. Reads were filtered to a minimum size of 1,500 base pairs (bp) and a maximum size of 3,500 bp to improve the quality of the output. This quality control was performed using a scalable protocol (https://artic-network.io/cov-2019-oncovid2019-bioinformatics-sop.pdf). Reads were aligned to the SARS- CoV-2 Wuhan-Hu-1 reference genome (GenBank sequence MN908947.3) using Minimap2 via ONT’s molecular biology and NGS pipeline. Variant calls were validated using Gatk4 v.0.11. Virological 2019 COVID-19 Lane Aligner was used to assign SARS-CoV-2 phylogenetic lineages.16 Complete SARS-CoV-2 genomes obtained in the lab were deposited to the GISAID database on January 31, 2021 (n=8731). Nextstrain’s NextClade software using default filtering and subsampling settings was used to perform phylogenetic analysis of US 20G clade isolates.17,18

AKNOWLEDGEMENT
This project was funded by Noblis and Tetracore. Special thank you to Tetracore for access to COVID-19 samples.

REFERENCES

© 2021 Noblis, Inc. All rights reserved. Proprietary to Noblis, Inc.