

First detection and report of SARS-CoV-2 Spike protein N501Y mutations in Oklahoma USA

Sai Narayanan¹, Girish Patil¹, Sunil More², Jeremiah Saliki¹, Anil Kaul³, Akhilesh Ramachandran¹

¹ Oklahoma Animal Disease Diagnostic Laboratory, College of Veterinary Medicine, Oklahoma State University, Stillwater OK 74078

² Department of Veterinary Pathobiology, College of Veterinary Medicine, Oklahoma State University, Stillwater, OK 74078

³ Center for Health Sciences, Oklahoma State University, Tulsa, OK 74107

Abstract:

We describe the detection of SARS-CoV-2 (VOC) B.1.1.7 lineage in Oklahoma, USA. Various mutations in the S gene and ORF8 with similarity to the genome of B.1.1.7 lineage were detected in 4 of the 6 genomes sequenced and reported here. The sequences have been made available in GISAID. Presence of novel lineages indicate the need for frequent whole genome sequencing to better understand pathogen dynamics in different geographical locations.

Keywords: SARS-CoV-2, B.1.1.7, Variant of Concern (VOC), mutations, Oklahoma (USA)

Since the start of the COVID-19 pandemic in December 2019, several lineages/variants of SARS-CoV-2 have been identified based on whole genome sequencing. As of January 21, 2021 approximately 420,000 whole genome sequences have been deposited in the GISAID database (<https://www.gisaid.org/>) (1,2). Recently a novel SARS-CoV-2 lineage (B.1.1.7.) was identified in Kent and Greater London in the United Kingdom (3) and is reported to have enhanced transmissibility. Infections by the B.1.1.7 lineage has since been reported from several other countries including the United States (4). We report the identification of B.1.1.7 in Oklahoma, USA.

COVID-19 was first reported in Oklahoma on March 6th 2020 (5). As of January 19th 2021, a total of 358,374 cases and 3,037 deaths have been registered in the state (6). Six clinical samples, from different counties in the state, collected between January 1st and 8th, 2021, were sequenced for this report. RNA was extracted using a commercial kit (MVP, ThermoFisher) and viral presence was detected by real-time PCR using TaqPath COVID-19 kit (Thermofisher, MA, USA) that targets the S gene, N gene and ORF 1ab. The S gene target of all the samples selected for sequencing either failed to amplify or had Ct values that were higher when compared to N gene and ORF1ab. Sequencing was performed using the MinION platform (Oxford Nanopore Technologies, UK) following the ARTIC protocol (7).

Sequenced genomes were assembled using canu (8). Reference assisted consensus assemblies were generated using minimap2 (9) and nanopolish (10), SARS-CoV-2 reference genome Wuhan-Hu-1 (GenBank ID: MN908947.3) was used for reference assembly. Individual genes were annotated using VIGOR (11). Annotated genes and the whole genomes were aligned using MUSCLE aligner in MEGA-X (12). The genome sequences were submitted to GISAID (1) and their GISAID classification and PANGO lineages (13) are shown in table 1. Four of the six samples sequenced carried SNPs similar to those described in the B.1.1.7 lineage (20I/501Y.V1 Variant of Concern (VOC) 202012/01) (4), including N501Y, 69/70 deletion, Y144 deletion, A570D, P681H, T716I, S982A, D1118H mutations. Q27Stop mutation in ORF 8 was detected in 3 out of these 4 samples. N501Y and the previously described D614G mutations (14,15) were detected in all 6 samples (Table 2).

Frequent whole genome sequencing is critical to monitor mutations in the viral genome. The evolution of highly transmissible strains of SARS-CoV-2 and its global spread underlines the need to remain vigilant and continue practicing necessary social measures to prevent further disease spread.

Genome	Coverage	GISAIID Accession No.	GISAIID Clade	PANGO Lineage
Oklahoma-ADDL-6	36932.1 ×	EPI_ISL_812346	GR	B.1.1.7
Oklahoma-ADDL-7	2080.4 ×	EPI_ISL_861422	G	B.1.1.7
Oklahoma-ADDL-8	1963.7 ×	EPI_ISL_861423	GR	B.1.1.7
Oklahoma-ADDL-9	2192.3 ×	EPI_ISL_861424	G	B.1.1.7
Oklahoma-ADDL-10	3992.5 ×	EPI_ISL_861425	GH	B.1
Oklahoma-ADDL-11	4399.6 ×	EPI_ISL_861426	GH	B.1

Table 1: Genome coverage of SARS-CoV-2 sequenced from different samples. GISAIID Accession No., Clade and PANGO (Phylogenetic Assignment of Named Global Outbreak) Lineage from GISAIID are shown.

Mutations observed (Nucleotide position)	Gene	Genomes with mutation
N501Y (A23063T)	S gene	Oklahoma-ADDL-6, 7,8,9,10,11
A570D (C23271A)	S gene	Oklahoma-ADDL-6,7,9
D614G (A23403G)	S gene	Oklahoma-ADDL-6, 7,8,9,10,11
P681H (C23604A)	S gene	Oklahoma-ADDL-6,7,8,9
T716I (C23709T)	S gene	Oklahoma-ADDL-6,7,8,9
S982A (T24506G)	S gene	Oklahoma-ADDL-6,7,8,9
D1118H (G24914C)	S gene	Oklahoma-ADDL-6
69/70 (21765 - 21770) DEL	S gene	Oklahoma-ADDL-6,7,8,9
Y144 DEL (21991 - 21993)	S gene	Oklahoma-ADDL-6,7,8,9
Q27Stop (C27972T)	ORF 8	Oklahoma-ADDL-6,7,8

Table 2: Selected mutations detected in S gene and ORF 8, and their respective amino acid changes when compared to NCBI reference genome (SARS-CoV-2, Wuhan Hu-1, NC 045512.2).

Ethics Statement:

This study was approved by the Institutional Review Board (Application number: IRB-20-357) at Oklahoma State University, Stillwater OK 74078, United States.

Acknowledgment

We thank Dr. Yiwei Wang and Dr. Jeremy Kaplan for their help in sample collection. Study was partially funded by FDA-CVM-VetLIRN (Grant no.5U18FD006671-02).

References:

1. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob challenges (Hoboken, NJ)* (2017) **1**:33–46. doi:10.1002/gch2.1018
2. Mak G. Important note: In the GISAID EpiFlu™ Database Access Agreement, you have accepted certain terms and conditions for viewing and using data regarding influenza viruses. To the extent the Database contains data relating to non-influenza viruses, the viewing. (2020)413498.
3. Chand, Meera; Hopkins, Susan; Dabrera, Gavin; Achison, Christina; Barclay, Wendy; Ferguson, Neil; Volz, Erik; Loman, Nick; Rambaut, Andrew; Barrett J (21 D 2020). I of novel S-C-2 variant: V of C 202012/01 (PDF) (Report). PHE p. 2. R 22 D 2020. Investigation of novel SARS-COV-2 variant- Variant of Concern 202012/01. (2020). Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/947048/Technical_Briefing_VOC_SH_NJL2_SH2.pdf
4. Emerging SARS-CoV-2 Variants. (2021) Available at: <https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/scientific-brief-emerging-variants.html>
5. First Case of Coronavirus Confirmed in Tulsa County. Available at: <https://www.kjrh.com/news/local-news/first-case-of-coronavirus-confirmed-in-tulsa-county>
6. OSDH-COVID-19 Tracker. (2020) Available at: https://coronavirus.health.ok.gov/sites/g/files/gmc786/f/2020.07.10_weekly_epi_report.pdf
7. Lu J, du Plessis L, Liu Z, Hill V, Kang M, Lin H, Sun J, François S, Kraemer MUG, Faria NR, et al. Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* (2020) **181**:997-1003.e9. doi:<https://doi.org/10.1016/j.cell.2020.04.023>
8. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* (2017) **27**:722–736. doi:10.1101/gr.215087.116
9. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* (2018) **34**:3094–3100. doi:10.1093/bioinformatics/bty191
10. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods* (2015) **12**:733–735. doi:10.1038/nmeth.3444
11. Wang S, Sundaram JP, Stockwell TB. VIGOR extended to annotate genomes for additional 12 different viruses. *Nucleic Acids Res* (2012) **40**:W186–W192. doi:10.1093/nar/gks528

12. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* (2018) **35**:1547–1549. doi:10.1093/molbev/msy096
13. Rambaut A, Holmes EC, O’Toole Á, Hill V, McCrone JT, Ruis C, du Plessis L, Pybus OG. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* (2020) **5**:1403–1407. doi:10.1038/s41564-020-0770-5
14. Narayanan S, Ritchey JC, Patil G, Teluguakula N, More S, Malayer J, Saliki J, Kaul A, Ramachandran A. SARS-CoV-2 genomes from Oklahoma, USA. *medRxiv* (2020)2020.09.15.20195420. doi:10.1101/2020.09.15.20195420
15. Daniloski Z, Guo X, Sanjana NE. The D614G mutation in SARS-CoV-2 Spike increases transduction of multiple human cell types. *bioRxiv* (2020)2020.06.14.151357. doi:10.1101/2020.06.14.151357