# Genomic Surveillance of SARS-CoV-2

**Ashley Nazario-Toole, Hui Xia, Holly Nguyen & Thomas F. Gibbons**

*59th Medical Wing Clinical Investigations & Research Support*

## Introduction

SARS-CoV-2, the virus that causes COVID-19, emerged in December 2019 & created a global pandemic resulting in nearly 3.5 million deaths worldwide. In the Department of Defense (DoD), over 290,000 personnel (civilians, dependents, and active duty) have been infected with the virus to date.[1]

In late 2020, new variants of SARS-CoV-2 emerged which contain genetic mutations affecting transmission, diagnostics, therapeutics, or immune escape. The CDC established 3 classifications for SARS-CoV-2 variants: Variants of Interest (VOI), Variants Concern (VOC), and Variants of High Consequence (VOHC). Circulating variants and novel viral mutations are monitored through whole-genome sequencing of clinical specimens. Routine genomic surveillance is particularly crucial for monitoring changes in the spike (S) protein, which plays a key role in host ACE2 receptor recognition and viral entry via cell membrane fusion.[2] Additionally, natural antibody response to SARS-CoV-2 is concentrated on two sections of the S protein, the N-terminal domain (NTD) and the receptor-binding domain (RBD).[3]

To enhance military readiness, the CIRS Laboratory developed a sequencing and bioinformatics workflow for molecular epidemiological SARS-CoV-2 surveillance using excess clinical specimens collected under IRB protocol FWH20200103E at Lackland AFB. This workflow includes viral RNA isolation, viral load quantification, tiling-based next-generation sequencing, bioinformatics analysis, and data visualization via phylogenetic trees and protein mapping.[4] Our findings demonstrate the utility and need for genomic surveillance in the DoD. Specifically, we successfully identified locally circulating VOCs (2 months prior to VOC designation) and have found additional, potentially significant mutations - such as a T478K RBD mutation that is also present in ~65% of cases reported in Mexico.

## Methods

**Excess Specimen**: Excess nasopharyngeal swabs were collected from JBSA beneficiaries & basic military trainees. Samples were diluted 1:1 in DNA/RNA shield.

**RNA extraction**: Viral RNA was extracted with Qiagen's EZ1 Virus Mini Kit. Viral load was quantified with a modified CDC RUO qRT-PCR assay.

**Library preparation**: PCR-amplicon (tiling) sequencing libraries were prepared using the CleanPlex SARS-CoV-2 Panel from Paragon Genomics.

**Illumina Sequencing**: Libraries were sequenced on Illumina MiSeq or NextSeq 500 systems at 2x151bp read length. Samples with a minimum of $1 \times 10^6$ reads and 20X genome coverage > 99.5% were analyzed.

**Sequencing analysis**: Illumina's DRAGEN Bio-IT platform aligned adapter trimmed FASTQ sequences to the SARS-CoV-2 reference genome (NC_045512) & generated variant call files (VCF) from primer trimmed alignment files. Bcftools was used to create consensus FASTAs from VCFs.

## Results

| Gene | Genome location | Nucleotide mutations | Amino acid substitutions |
|---|---|---|---|
| *ORF1ab* | 266 - 21,555 | 655 | 365 |
| *Spike (S)* | 21,563 - 25,384 | 136 | 90 |
| *ORF3a* | 25,393 – 26,220 | 61 | 44 |
| *Envelope (E)* | 26,245 – 26,472 | 2 | 1 |
| *Membrane(M)* | 26,523 – 27,191 | 17 | 2 |
| *ORF6* | 27,202 – 27,387 | 10 | 6 |
| *ORF7ab* | 27,394 - 27,887 | 21 | 14 |
| *ORF8* | 27,898 – 28,533 | 1 | 25 |
| *Nucleocapsid (N)* | 28,274 - 29,533 | 116 | 48 |

**Table 1: Number of SARS-CoV-2 nucleotide and amino acid mutations by gene.** From May 2020-May 2021, 284 excess NP swab specimens collected at JBSA-Lackland were sequenced using PCR-based tiling. A total of 1072 nucleotide substitutions were identified; causing 607 unique amino acid substitutions.

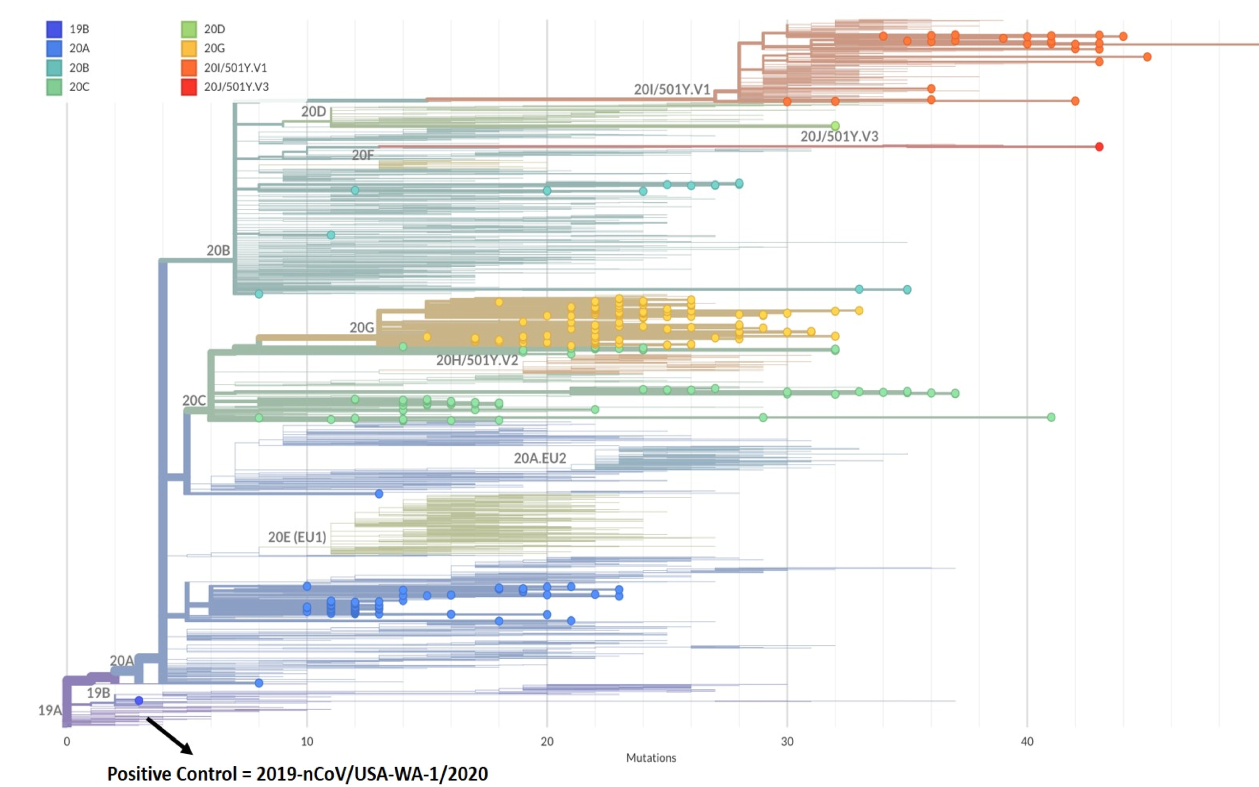


**Figure 1: Genomic and epidemiological diversity of SARS-CoV-2 sequences.**
59MDW SAR-CoV-2 sequences were uploaded to the Global Influenza and Infectious Disease (GISAID) online phylogenetic tool, Nextclade, to assign lineages and examine JBSA specimen sequence diversity relative to world-wide SARS-CoV-2 viruses.

| PANGO Lineage | # | Notable RBD mutations | CDC Classification |
|---|---|---|---|
| B.1 | 49 | | |
| B.1.2 | 97 | | |
| B.1.1.222 | 6 | T478K | |
| **B.1.1.7** | **30** | N501Y | **VOC** |
| **P.1** | **1** | K417T, E484K, N501Y | **VOC** |
| **B.1.351** | **0** | K417T, E484K, N501Y | **VOC** |
| **B.1.429** | **15** | L452R | **VOC** |
| Others | 86 | | |
| **Total** | **284** | | |

**Table 2: 59MDW SARS-CoV-2 by Pangolin Lineage & CDC classification.** Of 284 sequenced specimens, 46 VOCs have been identified:

***B.1.429 (USA)** detected in a **2 Jan 2021** specimen, two months before VOC classification.

***B.1.1.7 (UK)** first detected in **18 Feb 2021** specimen.

***P.1 (Brazil/Japan)** was detected in a specimen collected on **16 Apr 2021.**

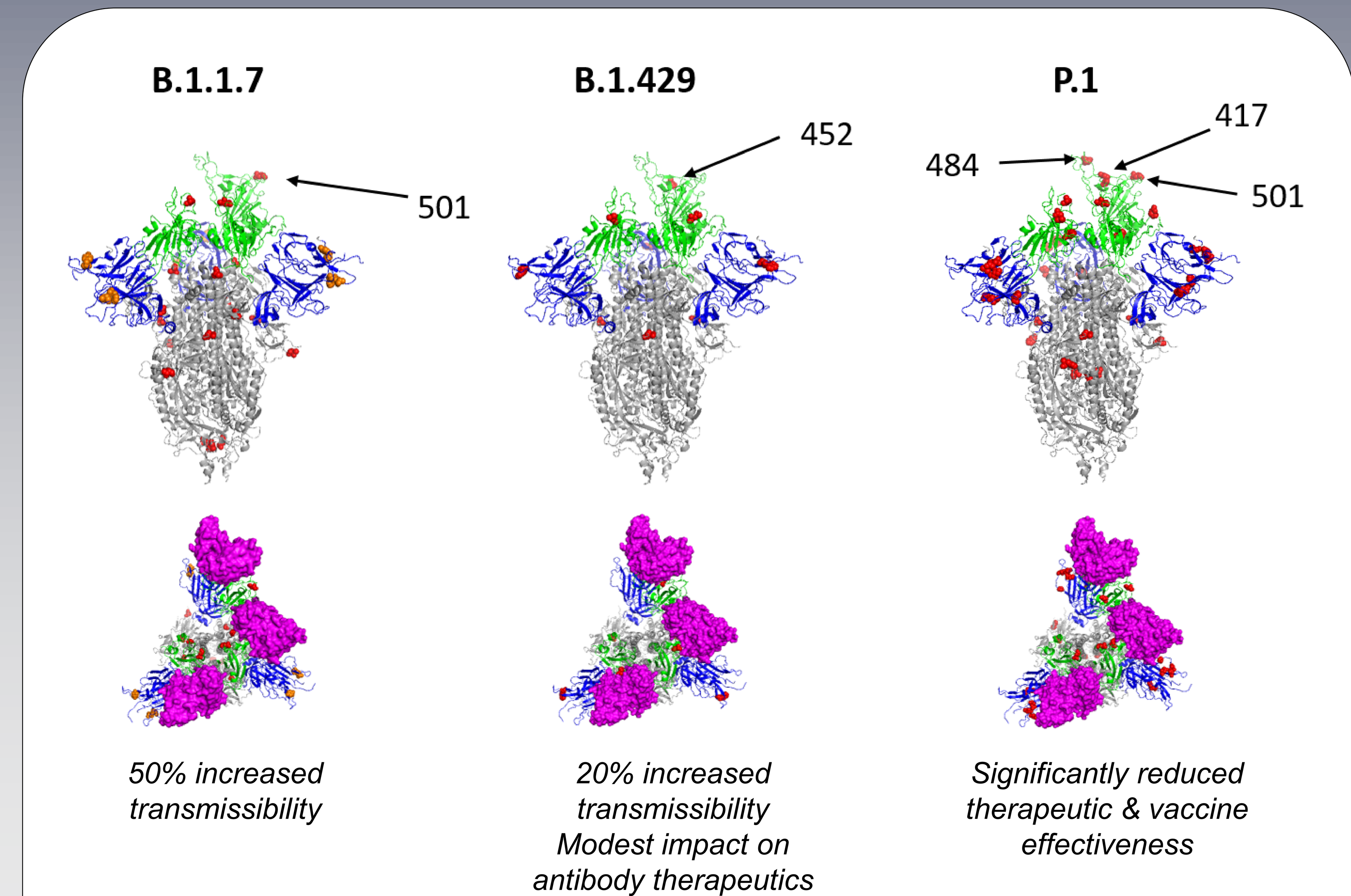


**Figure 2: 59MDW VOC Spike protein mutations.** Top: Red = Amino acid substitutions. Green = Receptor Binding Domain (RBD). Bottom: Visualization of receptor binding domain & neutralizing antibodies interactions.

## Conclusions

- Viral mutations and variants in the United States are routinely monitored through sequence-based surveillance, laboratory studies, and epidemiological investigations.
- CIRS developed and implemented a method for robust genomic surveillance of SARS-CoV-2 to identify and report genomic changes that are of Public Health interest.
- Viral genome sequences and the associated sequence quality data can be shared as FASTA, VCF & genomic coverage files.
- The rapid identification and modeling of spike protein mutations provide valuable insight for vaccine efficacy.

## Bibliography

1. US Department of Defense Coronavirus Response Dashboard: https://www.defense.gov/explore/spotlight/coronavirus (updated 20 May 2021)
2. Williams, Thomas C. *et al.* (2021). SARS-CoV-2 evolution and vaccines: cause for concern? The Lancet Respiratory Medicine, Volume 9, Issue 4, p 333 – 335.
3. Greaney *et al.* Comprehensive mapping of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human serum antibodies. Cell Host & Microbe doi: 10.1016/j.chom.2021.02.003.
4. Nazario-Toole, A. E., Xia, H., Gibbons, T.F. (2021) Whole-genome Sequencing of SARS-CoV-2: Using Phylogeny & Structural Modeling to Contextualize Local Viral Evolution. Military Medicine. usab031.

The views expressed are those of the authors and do not reflect the official views or policy of the Department of Defense or its Components. The authors do not have any financial interest in the companies whose materials are discussed in this poster.