

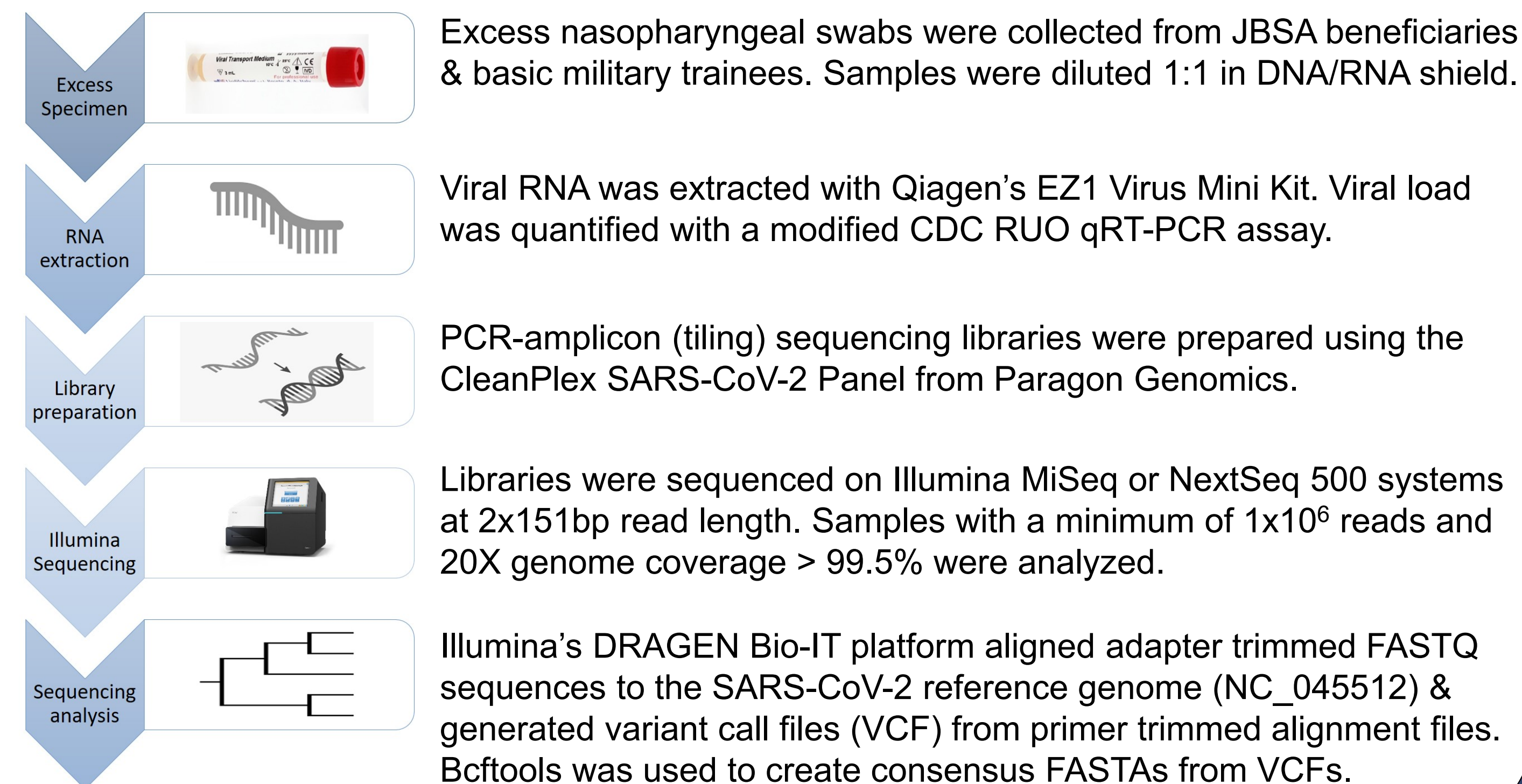
Introduction

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



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.

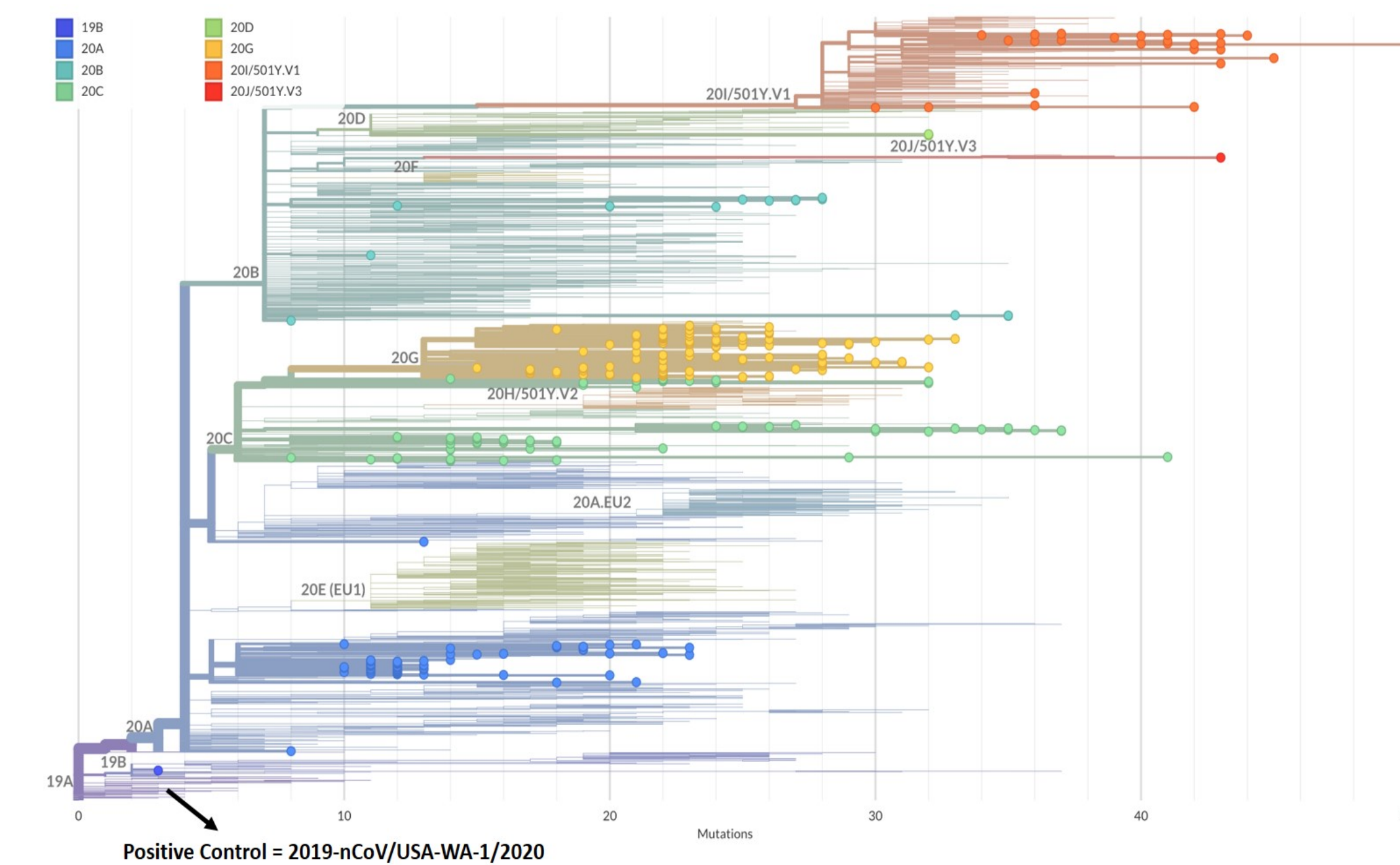


Figure 1: Genomic and epidemiological diversity of SARS-CoV-2 sequences.

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**.

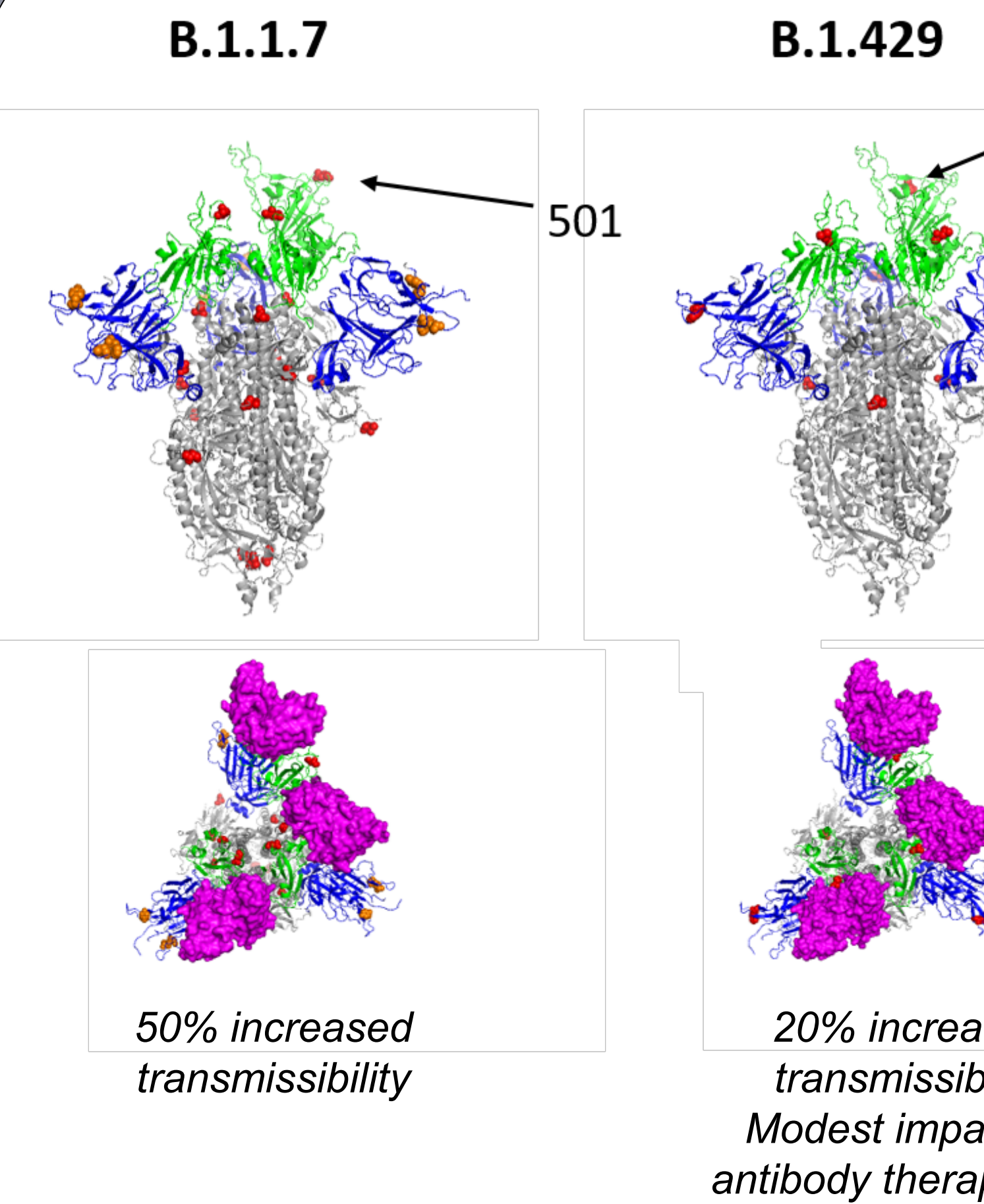


Figure 2: 59MDW VOC Spike protein mutations. Green = Receptor Binding Domain (RBD) & neutralizing antibody binding sites. Red = 501Y mutation site. Substitutions are shown in green. B.1.1.7 shows 50% increased transmissibility. B.1.429 shows 20% increased transmissibility and a modest impact on antibody therapy.

Conclusion

- Viral mutations and variants in the spike protein are monitored through sequence-based surveillance, epidemiological investigations, and epidemiological studies.
- CIRS developed and implemented a genomic surveillance of SARS-CoV-2 to identify mutations that are of Public Health importance.
- Viral genome sequences and the associated epidemiological data can be shared as FASTA, VCF, and other formats.
- The rapid identification and modeling of these variants provide valuable insight for vaccine development.

Bibliography

1. US Department of Defense COVID-19. <https://www.defense.gov/explore/spotlight> (2021)
2. Williams, Thomas C. *et al.* (2021). SARS-CoV-2: A cause for concern? *The Lancet Respiratory Medicine* 333 – 335.
3. Greaney *et al.* Comprehensive mapping of receptor-binding domain that affect receptor binding and antibody neutralization. *Cell Host & Microbe* doi: 10.1016/j.chom.2021.05.004
4. Nazario-Toole, A. E., Xia, H., Giblin, J. *et al.* Sequencing of SARS-CoV-2: Using Phylogenetic Contextualize Local Viral Evolution. *Military Medicine* 186(10):e19111 (2021)