Next Generation Sequencing: Monitoring Change during the COVID-19 Pandemic

Since January 2020 nations and communities across the globe have battled the COVID-19 pandemic. To date over 31 million cases of COVID-19 have been reported with more than 560,000 deaths recorded in the United States alone. In December 2020 the first COVID-19 vaccines were rolled-out in the U.S. During the first quarter of 2021 these became more available to the general public due to increased production and an increase in the range of approved alternatives. In the U.S. nearly 200 million vaccine doses have been administered thus far; a remarkable accomplishment. Although daily case numbers in the U.S and in many other nations have trended downward since peaking in January 2021 a new concern has arisen amidst the pandemic. Public health agencies across the globe have begun identifying genetic variants of the SARS-CoV-2 virus. While some variants cause limited additional concern, other strains have demonstrated increased transmissibility, resistance to treatment, and the capability to elude immunity thought to be provided through recovery from previous infection or vaccination. At this time, the CDC recognizes five "variants of concern" (VOC) due to increased risk for infection and the ability to cause greater harm to those that are exposed (B.1.1.7, P.1, B.1.351, B.1.427, B.1.429).[1] Although there has been significant innovation over the last twelve months in the development and implementation of diagnostic testing, next generation sequencing (NGS) will be a cornerstone in monitoring the spread of these variants of concern.

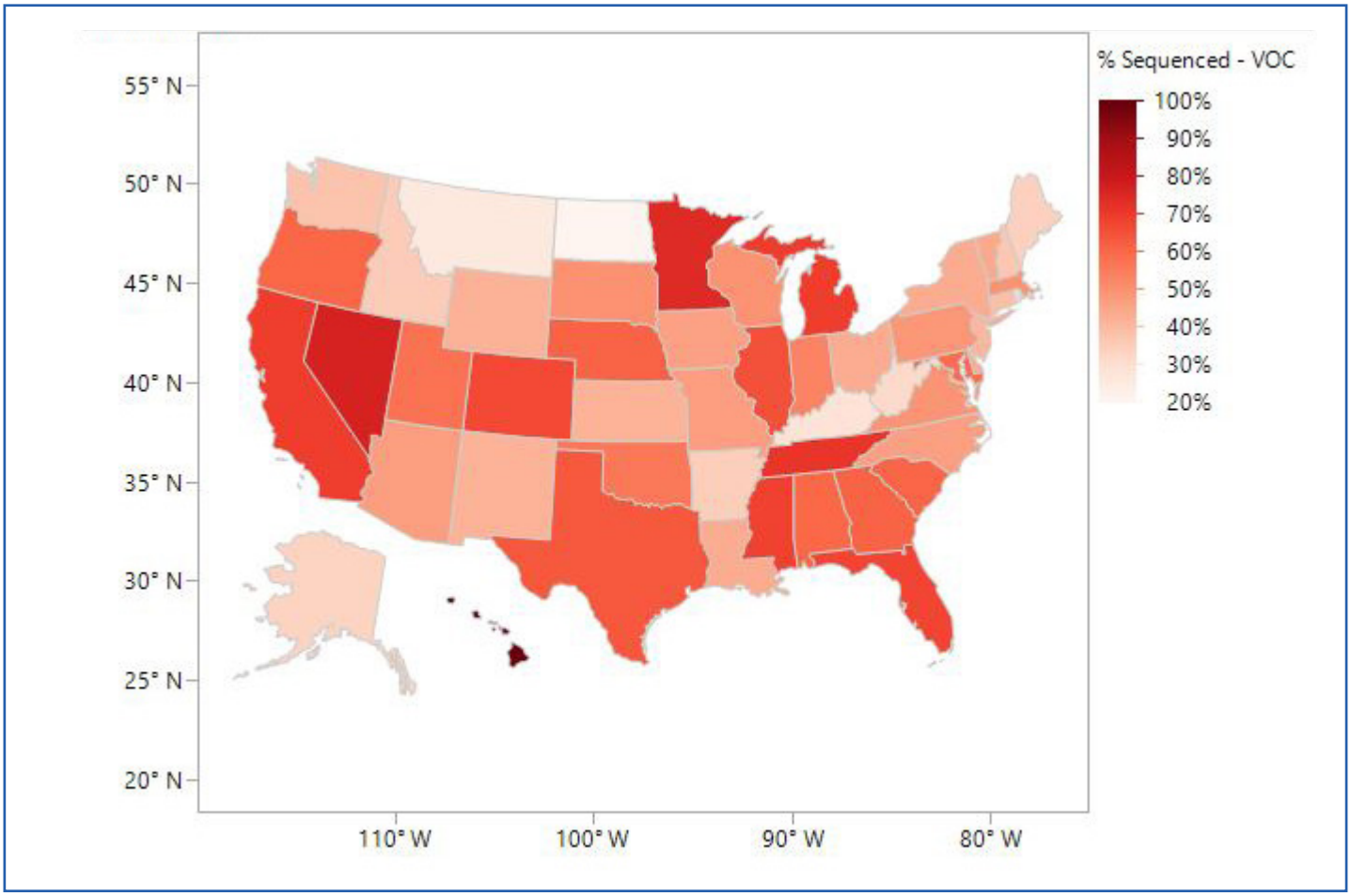
Massively parallel sequencing, or next-generation sequencing (NGS) is a molecular technique that has revolutionized the sequencing industry over the last 20+ years. NGS employs amplicon-based sequencing by synthesis technology to make thousands of copies of a segment of viral DNA. During each sequencing cycle, a single deoxynucleoside triphosphate(dNTP) is added to the nucleic acid chain and a fluorescent dye is imaged to identify the base and then cleaved to allow incorporation of the next nucleotide. During the primary analysis and datafile generation, base calls are made directly from signal intensity measurements during each sequencing cycle, resulting in base-by-base sequencing. Following data generation, bioinformatic tools are utilized to perform alignment of the read sequences to create the entire viral genome in silico. Aegis utilizes the Illumina® COVIDSeq Test performed on the NovaSeq 6000 platform. The Illumina® DRAGEN COVID Lineage App uses the FASTQ files to align reads to a SARS-CoV-2 reference genome and reports coverage of targeted regions. It performs Kmer-based detection and then performs Map/Align, Variant Calling, Consensus Sequence generation, and lineage/clade analysis using Pangolin and NextClade for samples with at least 90 out of 98 SARS-CoV-2 targets detected. Next-generation sequencing allows for high-throughput large scale data generation and identification of variant strains of SARS-CoV-2 and it is integral to the national effort to slow the spread of the virus.

As of April 2020, Aegis has completed diagnostic testing utilizing real-time reverse transcriptase polymerase chain reaction testing to assist in determination of the presence of COVID-19 in over 4.5 million samples provided nationwide. Aegis has entered into a collaborative agreement with the CDC with a sequencing target of 10,000 patient samples per week providing a significant boost to surveillance efforts within the United States. At the time this document was prepared, over 20,000 positive samples had been analyzed utilizing the next generation sequencing technique in our laboratory (Table 1). Our analysis has identified significant transmission of the variants of concern throughout the United States (Figure 1).

Table 1: Number of Samples Sequenced per State

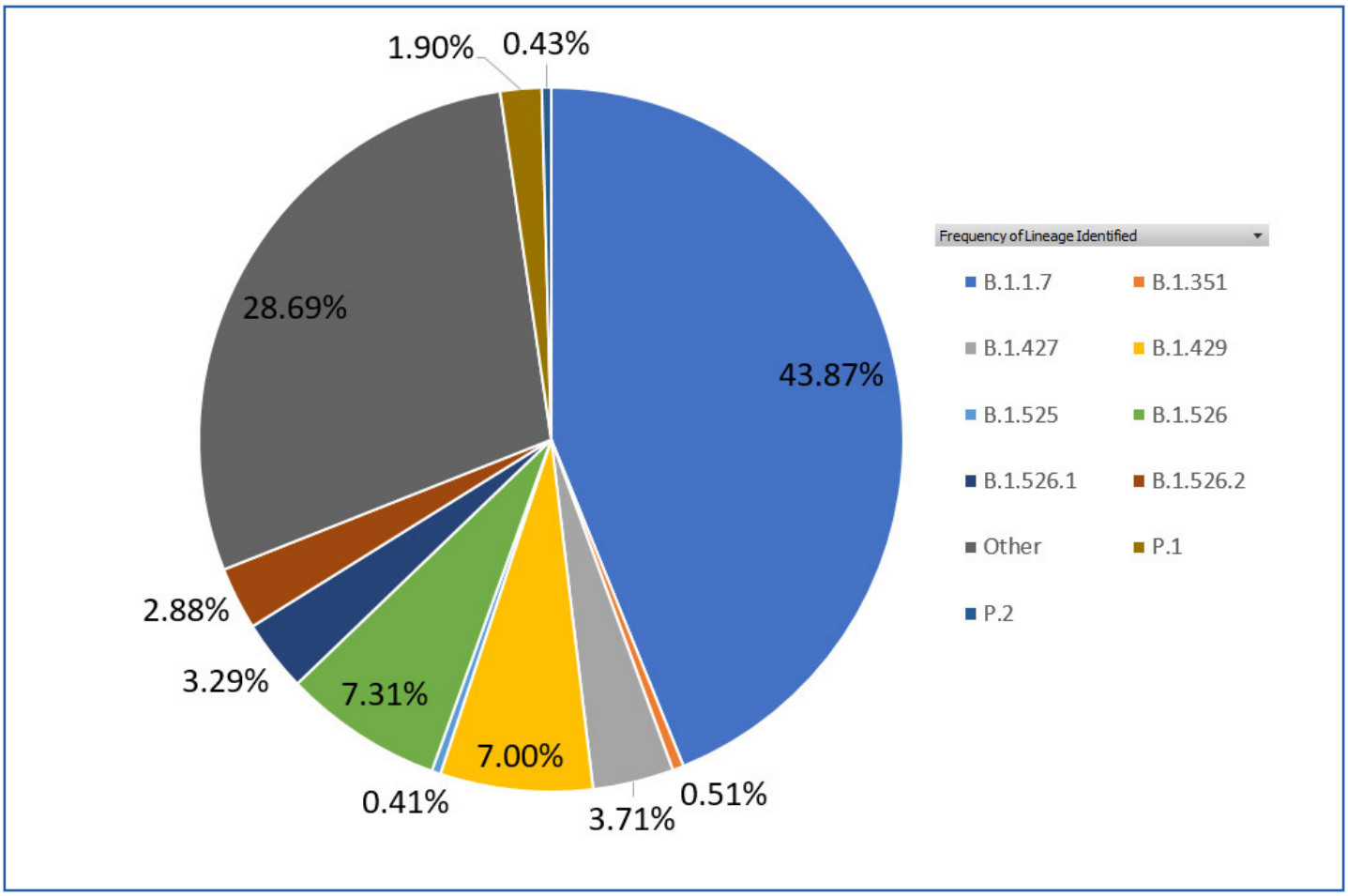
State	Samples Sequenced	State	Samples Sequenced	State	Samples Sequenced
Alabama	156	Louisiana	101	Oklahoma	62
Alaska	12	Maine	53	Oregon	190
Arizona	160	Maryland	566	Pennsylvania	700
Arkansas	26	Massachusetts	851	Puerto Rico	90
California	1269	Michigan	2171	Rhode Island	341
Colorado	629	Minnesota	557	South Carolina	135
Connecticut	137	Mississippi	38	South Dakota	28
Delaware	76	Missouri	246	Tennessee	1004
District of Columbia	9	Montana	39	Texas	984
Florida	1785	Nebraska	114	Utah	178
Georgia	411	Nevada	48	Vermont	45
Hawaii	8	New Hampshire	49	Virginia	189
Idaho	116	New Jersey	2384	Washington	16
Illinois	1287	New Mexico	50	West Virginia	268
Indiana	253	New York	828	Wisconsin	356
Iowa	209	North Carolina	439	Wyoming	19
Kansas	50	North Dakota	5		
Kentucky	154	Ohio	360		

Figure 1: Frequency of Variant of Concern Identification by State



Amongst the samples that have undergone next generation sequencing in our laboratory, over 50% have been determined to contain VOC (Figure 2). The predominant VOC identified to date has been of the B.1.1.7. lineage, which was first identified in the United Kingdom in September 2020. This mutant strain has demonstrated greater than 50% increased transmissibility in comparison to earlier variants and is associated with increased morbidity and mortality in those that contract the virus.1 Though early evidence suggests limited impact of B.1.1.7. on the effectiveness of currently available vaccines and therapeutics, other studies have demonstrated that some variants carry an additional mutation that may impact immune response.[2-3] While currently not common as compared to B.1.1.7. close monitoring of the spread of P.1 (4.2%) and B.1.351(0.5%) lineages is warranted as both have proven evasive to both vaccine-induced immunity and therapeutics.

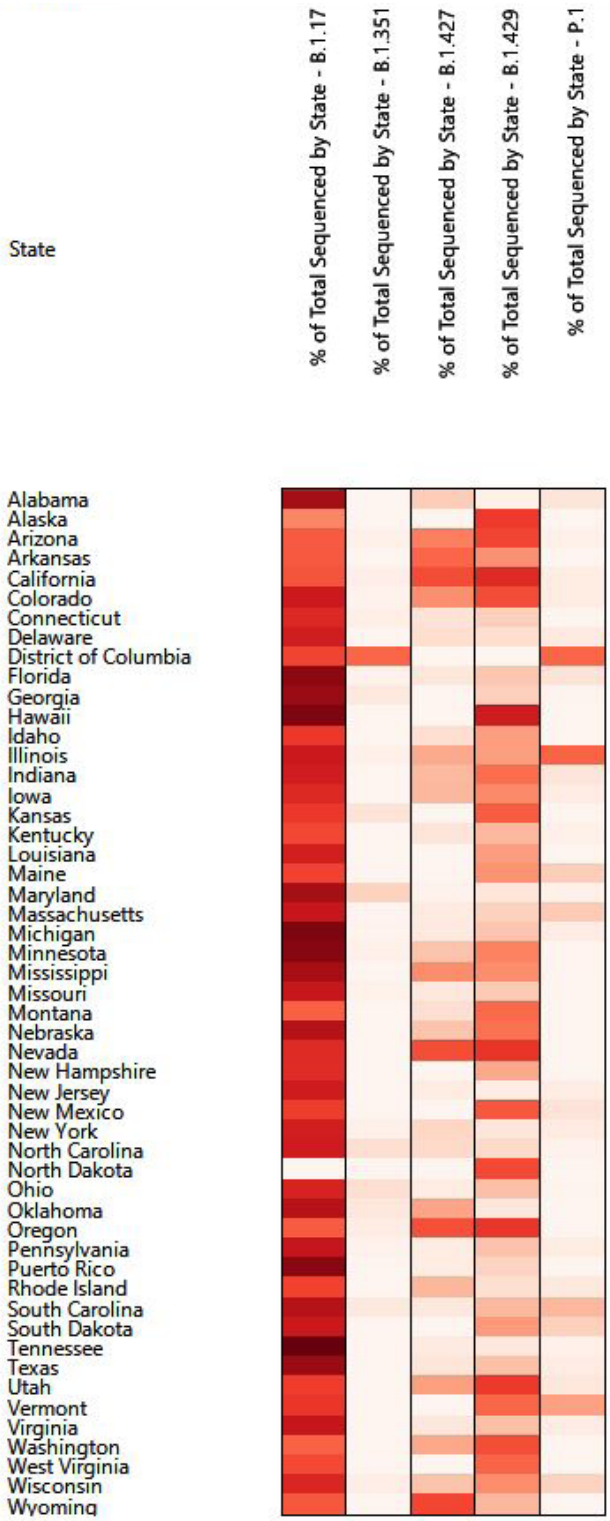
Figure 2: Comparison of Identified Variants (Samples Collected from 2/14/21-4/13/21)



There appears to be relatively significant differences in terms of regional spread of the different variants (Figure 3). The heat map below displays pockets of infection involving particular variants. Although B.1.1.7 is the predominant variant identified thus far, two other variants believed to have originated in California, B.1.427 and B.1.429, have begun to spread throughout the western part of the country. At the time that this document was prepared, 11 states had demonstrated higher prevalence of B.1.427 or B.1.429 than B.1.1.7. Despite being thought to be slightly less transmissible than B.1.1.7, B.1.427 and B.1.429 have both shown a reduced effectiveness of medications used to treat those with COVID-19, as well as the possibility of reducing immunity provided by vaccine.

As more is learned regarding the ability of variants to negatively impact current therapeutics and evade immunity gained through previous infection of immunization, it will be key to continue to monitor their transmissibility and spread throughout the county. Armed with this information the general public can be advised regarding appropriate precautions and Public Health Officials and Healthcare resources can be effectively mobilized to limit spread of harmful variants. Next generation sequencing is a powerful tool that is capable of efficiently identifying the presence of genetic mutations after initial positive diagnosis has occurred, and it will be of paramount importance in continuing the fight against the pandemic.

Figure 3: Heatmap of Variants of Concern by State (Dark Red = Higher Frequency VOC as State level)



References:
1. SARS-CoV-2 Variants of Concern. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html. Accessed April 20, 2021.
2. Science Brief: Emerging SARS-CoV-2 Variants. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/scientific-brief-emerging-variants.html. Accessed April 20, 2021.
3. Wise J. Covid-19: The E484K mutation and the risks it poses. The BMJ. https://www.bmj.com/content/372/bmj.n359. Published February 5, 2021. Accessed April 20, 2021.

