

**CFSAN/OAO**  
**BIostatISTICS AND BIOinformatics STAFF****WASTEWATER SARS-COV2 ANALYSIS REPORT**

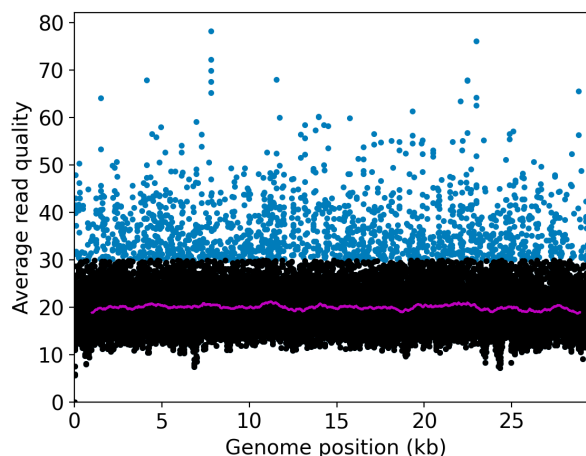
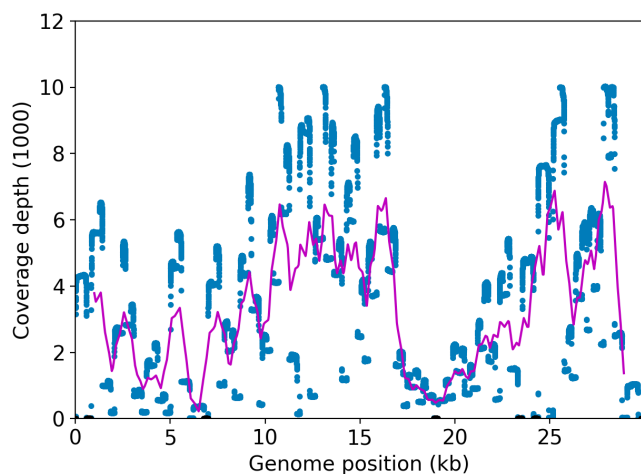
Sample name:	SRR18910148
Date generated:	2022-06-03, 15:19:16 UTC
Timestamp of C-WAP version used:	Sat May 28 00:55:30 2022 +0000
Executed by:	tunc kayikcioglu ( <a href="mailto:tunc.kayikcioglu@fda.hhs.gov">tunc.kayikcioglu@fda.hhs.gov</a> )
Executed on:	172.20.17.36 (aka compute-dy-r54xlarge-8)

**Sequencing summary**

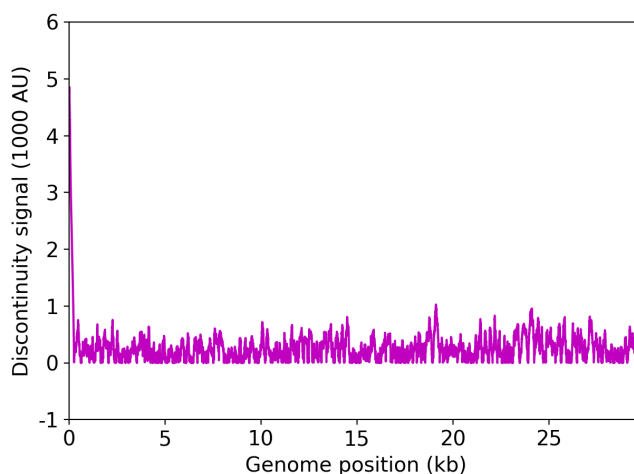
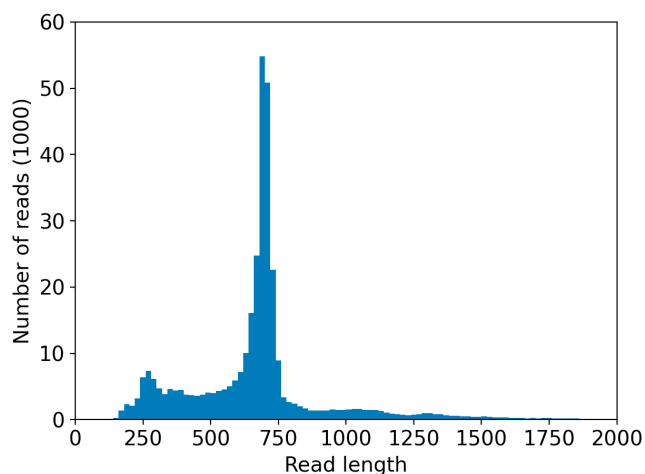
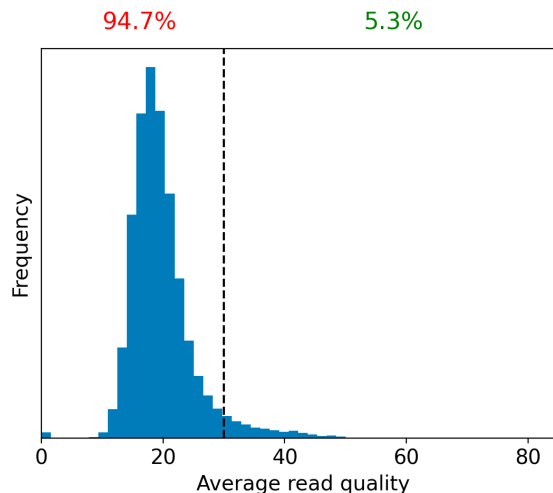
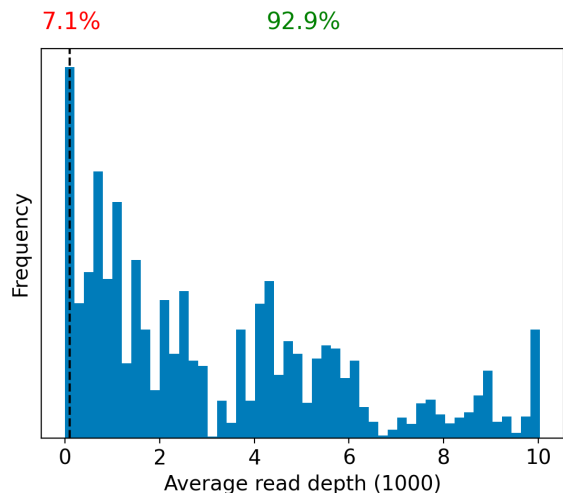
Sequencing chemistry:	WGS with MinION
Source site:	<a href="#">USA: Maryland (missing.?)</a>
Sampling date:	2022-03-16
Collected by:	FDA Center for Food Safety and Applied Nutrition
Sequenced by:	Missing
Total number of reads:	344810
Reads aligned:	199154 (57%)
Average read quality:	19.1
Average read length:	679
Reads passing filter:	199154 (57%)
Average read quality passing filter:	19.1
Average read length passing filter:	679
Average coverage passing filter:	4522X

A read passes filter if the read length after adaptor trimming  $\geq 30$  and minimum read quality  $\geq 20$  within a sliding window of width 4.

**Overall sequence characteristics**



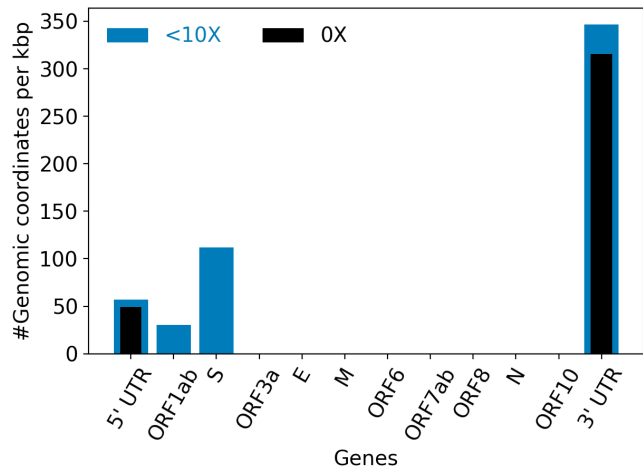
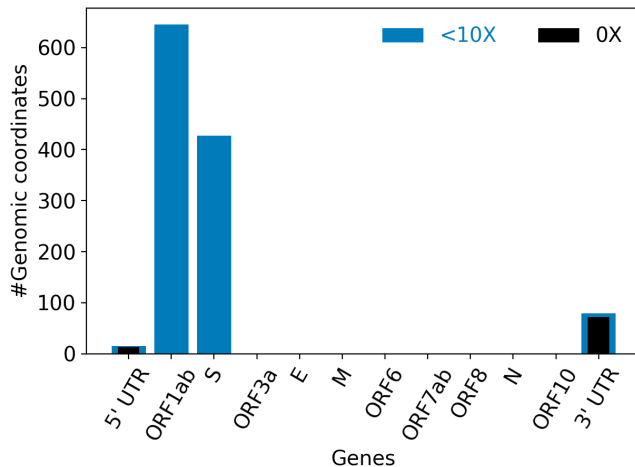
NOTE: The red shaded areas marked with a (\*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.



	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	-1nt (0%)	-1nt (0%)

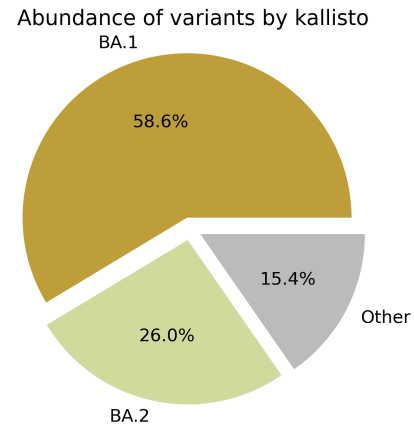
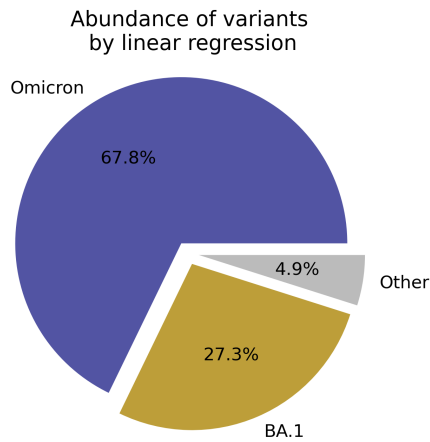
All genomic coordinates:	85nt (0%)	1166nt (3%)
Common SNPs:	0nt (0%)	2nt (5%)
Diverse SNPs:	27nt (5%)	28nt (5%)
Rare SNPs:	14nt (1%)	18nt (1%)

SNPs refer to the polymorphic sites currently in circulation that were detected out of recent GISAID entries. The sites that differ from the SC2 reference sequence are denoted as "common" if [90%, 100%] of the submissions carry this mutation, whereas those that are prevalent in [0%,10%] of the submissions are grouped under the "rare" category. The population is still diverse at the mutation sites that are observed in (10%,90%) of the entries and these coordinates are grouped under the "diverse" category.



Hits to SARS-Cov2 genome (kraken2):	196899 reads (57.10%)
Hits to human genome (kraken2):	2480 reads (0.72%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
Most abundant organisms (kraken2, family level):	Coronaviridae (57.10%) Campylobacteraceae (19.01%) Pseudomonadaceae (3.02%) Flavobacteriaceae (1.46%) Comamonadaceae (1.45%)

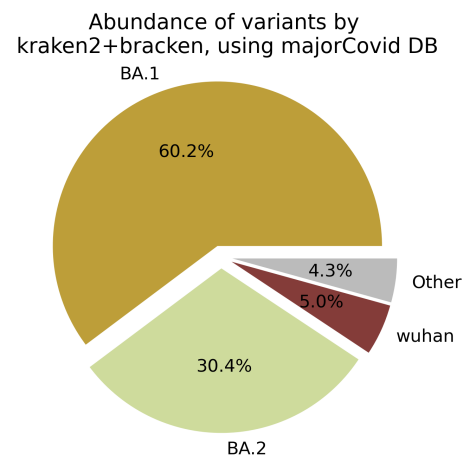
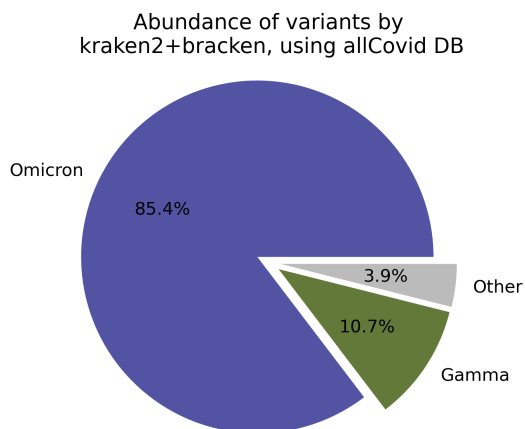
## Detected variants (Experimental)



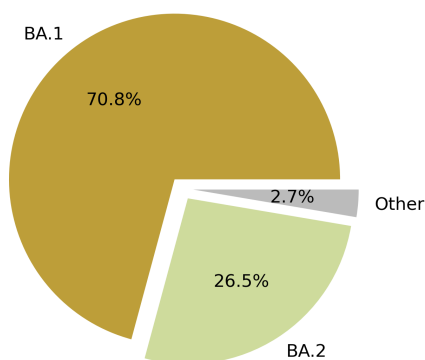
Based on deconvolution, [B.1.1.529](#) is estimated to constitute 67.80% of the viral particles and hence is the most abundant variant in the sample. The  $R^2$  for the linear regression was 0.54. Variants that were detected less than 5% were grouped under "Other"

Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [Unassigned](#) lineage. If this is a sample consisting of a single source of pathogens or an overwhelming majority of the different sources are infected with the same variant, the sample is dominated by this variant.

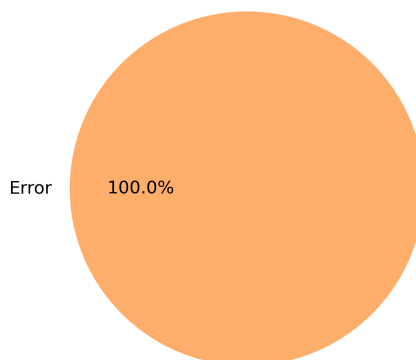
Based on mapping individual reads to the variant consensus sequences in the reference database, kallisto predicts that the sample is dominated by [BA.1](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.



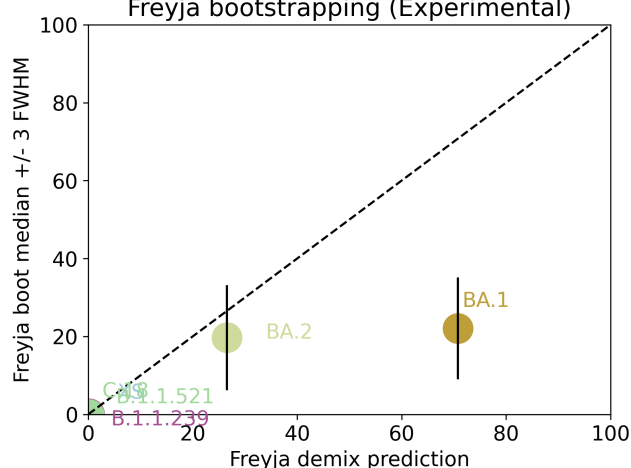
Abundance of variants by Freyja



Abundance of variants by LCS



Freyja bootstrapping (Experimental)



Under the assumption that the presence of a variant requires the detection of all respective mutations of the variant, the characteristic mutations which support the presence of the respective variant are indicated in the respective column of the table. Numbers show the number of mutations detected, if any, and the number of mutations expected to be present based on the variant definitions.

VOC	<a href="#">B.1.617.2</a>	<a href="#">BA.1</a>	<a href="#">BA.2</a>	<a href="#">BA.3</a>	<a href="#">BA.4</a>	<a href="#">BA.5</a>
Characteristic mutations detected	(2 of 13) S:G142D S:T478K	(15 of 26) M:D3G NUC:C15240T NUC:C25584T NUC:T13195C ORF1AB:A2710T ORF1AB:I3758V ORF1AB:K856R S:A67V S:G446S S:G496S S:L981F S:N856K S:Q493R S:T547K S:T95I	(13 of 31) NUC:C12880T NUC:C15714T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:L3201F ORF1AB:T3090I ORF1AB:T842I S:Q493R S:S371F S:T19I	(10 of 21) NUC:C12880T NUC:C15714T NUC:C26858T NUC:G10447A ORF1AB:G1307S ORF1AB:T3090I S:A67V S:G446S S:Q493R S:S371F	(12 of 31) NUC:C12880T NUC:C15714T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:T3090I ORF1AB:T842I S:S371F S:T19I S:V213G	(11 of 28) NUC:C12880T NUC:C15714T NUC:C25584T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:T3090I ORF1AB:T842I S:S371F S:T19I S:V213G

## Detected mutations

Excluded from this pdf version due to file size limitations.