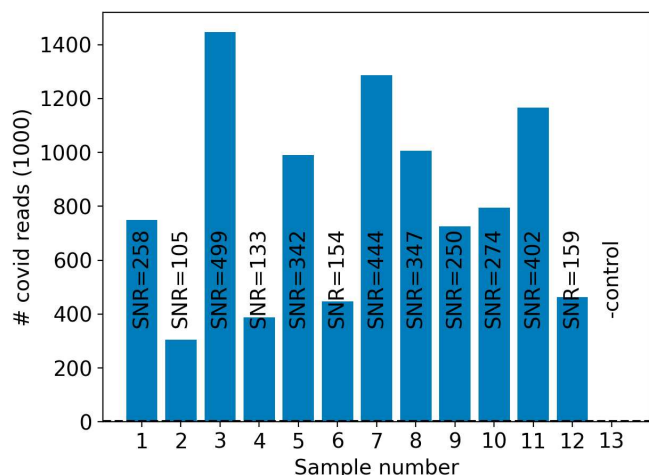


**CFSAN/OAO
BIostatistics and Bioinformatics Staff**

WASTEWATER SARS-COV2 ANALYSIS REPORT

Summary

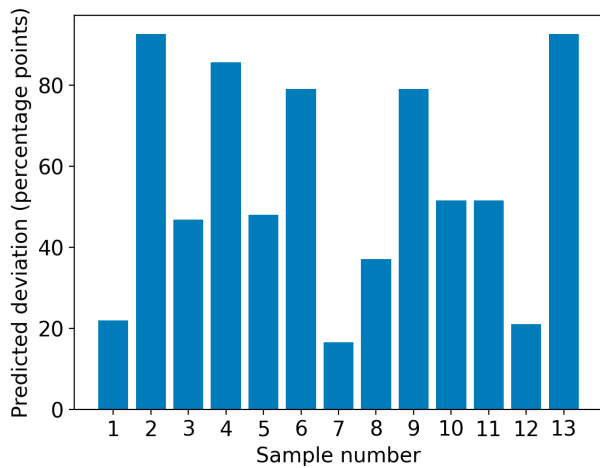
Sample#	Sample name	Total #reads	Reads aligned PF*	Genomic coordinates 0X	Genomic coordinates <10X
1	CFSANSMP000119220	1240436	748699 (60%)	401nt (1%)	1014nt (3%)
2	CFSANSMP000119221	420127	304556 (72%)	481nt (1%)	3391nt (11%)
3	CFSANSMP000119222	1536416	1447541 (94%)	406nt (1%)	733nt (2%)
4	CFSANSMP000119223	487373	387336 (79%)	509nt (1%)	548nt (1%)
5	CFSANSMP000119224	1115551	990602 (88%)	410nt (1%)	1350nt (4%)
6	CFSANSMP000119225	561738	448028 (79%)	511nt (1%)	1874nt (6%)
7	CFSANSMP000119226	1638438	1286629 (78%)	475nt (1%)	521nt (1%)
8	CFSANSMP000119227	1336574	1005993 (75%)	480nt (1%)	513nt (1%)
9	CFSANSMP000119228	1037901	725752 (69%)	503nt (1%)	517nt (1%)
10	CFSANSMP000119229	1172436	794585 (67%)	441nt (1%)	511nt (1%)
11	CFSANSMP000119231	1298984	1166947 (89%)	477nt (1%)	512nt (1%)
12	<u>unclassified</u>	569329	462282 (81%)	406nt (1%)	509nt (1%)
13	<u>Water</u>	2964	2896 (97%)	13665nt (45%)	17534nt (58%)



*Quantity of raw reads that align to the reference sequence and pass filter, i.e. the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4. SNR refers to the ratio of SC2-mapping reads aligned that pass filter in the sample vs. that in the auto-detected negative control samples (if any). The dashed line represents the baseline level of covid reads detected from the negative control or their average if multiple negative controls were included.

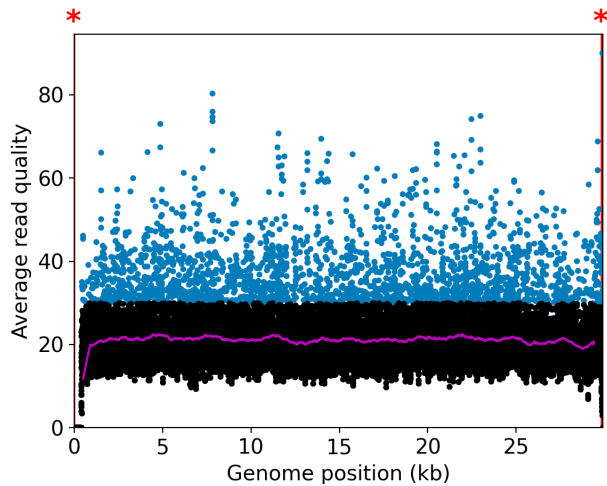
QC category	Subjective definition	Objective metrics
A	No QC issues evident	0x coordinates <1% 10x coordinates <5% average coverage > 1000X average quality score >35 for Illumina, >15 if ONT, >70 if PacBio HiFi most abundant taxon is coronavirinae
B	Some QC issues, but accurate variant calling possible	0x coordinates <20% 10X coordinates < 40% >80% of diverse SNPs covered average coverage > 100X average quality score >35 for Illumina >15 if ONT, >70 if PacBio HiFi
C	Some QC issues, and accurate variant calling impossible	0x coordinates <99% 10X coordinates <95%
F	Significant QC/study design issues	Contamination (SNR<50) No/negligible coverage (< 1X) Biological/technical replicates' results are irreconcilable.

Sample Number	Suggested category	Suggested QC flags
1	B/C	low_coverage_breadth
2	B/C	low_coverage_breadth
3	B/C	low_coverage_breadth
4	B/C	low_coverage_breadth
5	B/C	low_coverage_breadth
6	B/C	low_coverage_breadth
7	B/C	low_coverage_breadth
8	B/C	low_coverage_breadth
9	B/C	low_coverage_breadth
10	B/C	low_coverage_breadth
11	B/C	low_coverage_breadth
12	B/C	low_coverage_breadth
13	F	sample_contamination

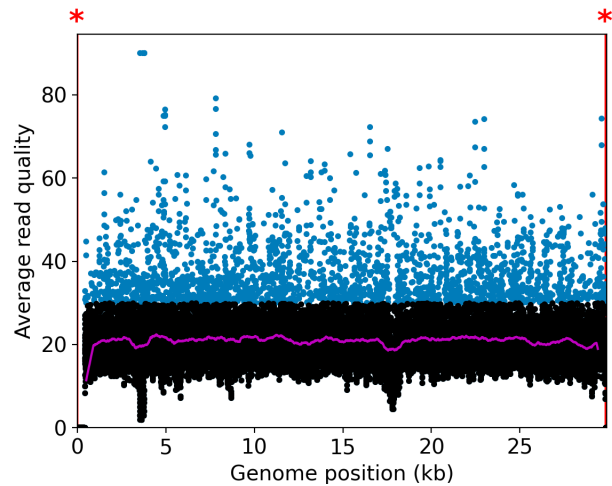


Machine-learning based prediction of the SC2 variant calling accuracy of Freyja of this dataset. The model is a random forest trained on FDA/CFSAN's experimental wastewater WGS data obtained in January 2022 and aims to assess the impact of the potential coverage gaps on the variant abundance estimates. The plotted values represent the predicted deviation of the omicron percentage points from the value that would have been obtained if the coverage was near-complete.

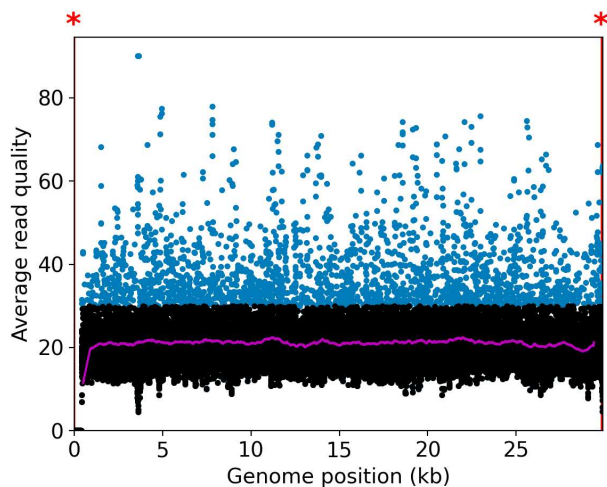
[CFSANSMP000119220](#)



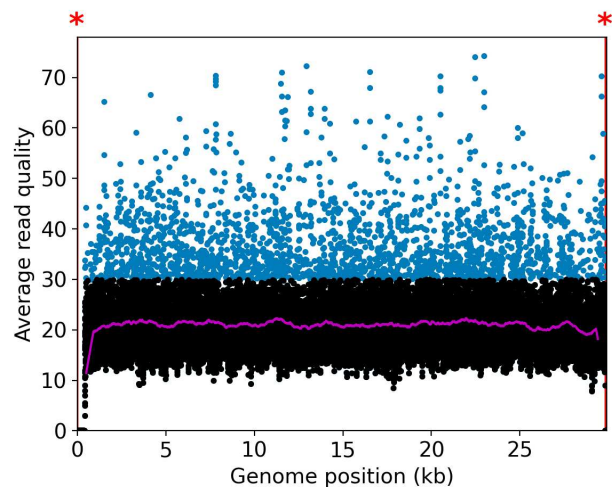
[CFSANSMP000119221](#)



[CFSANSMP000119222](#)



[CFSANSMP000119223](#)

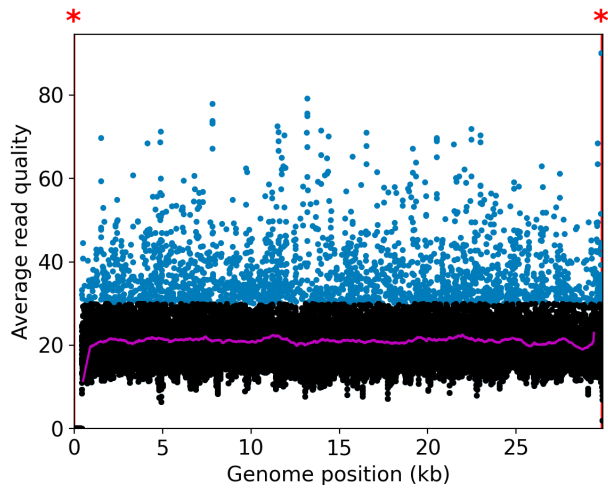


[CFSANSMP000119224](#)

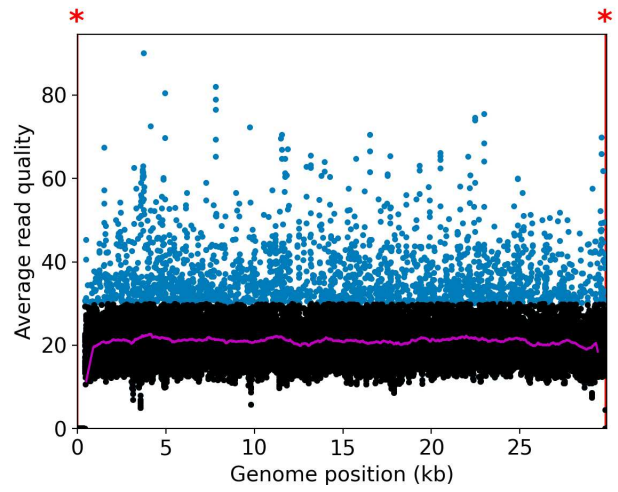


[CFSANSMP000119225](#)

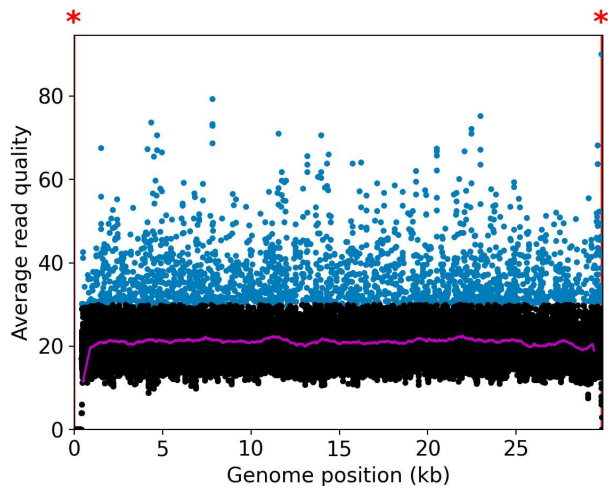




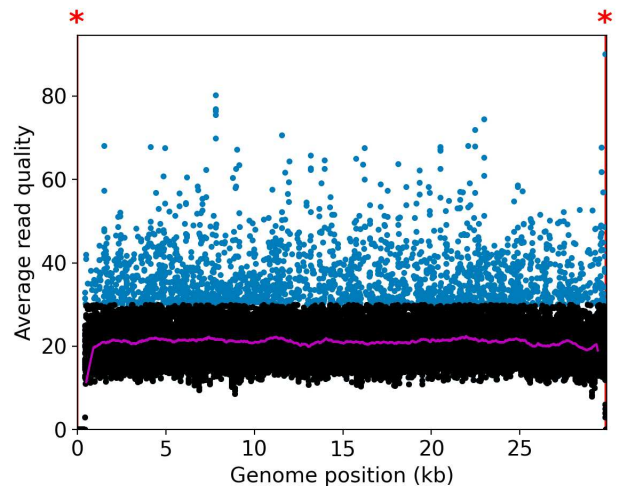
[CFSANSMP000119226](#)



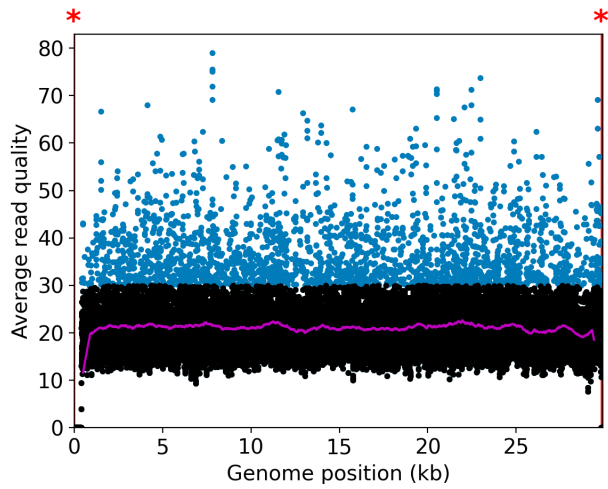
[CFSANSMP000119227](#)



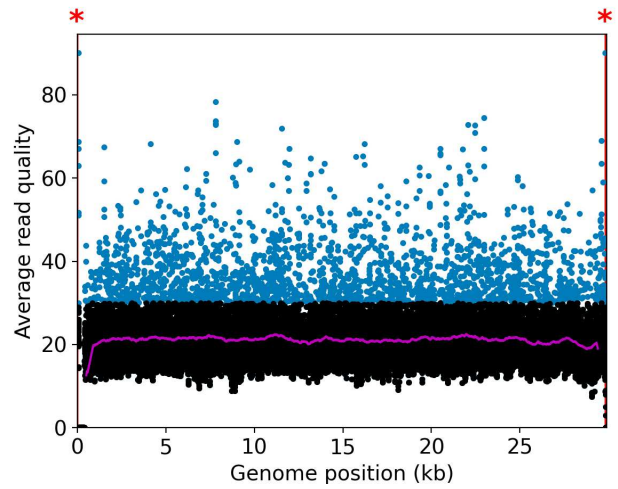
[CFSANSMP000119228](#)



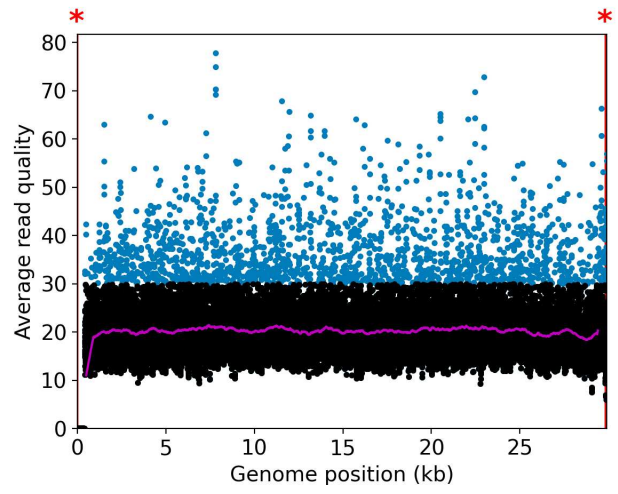
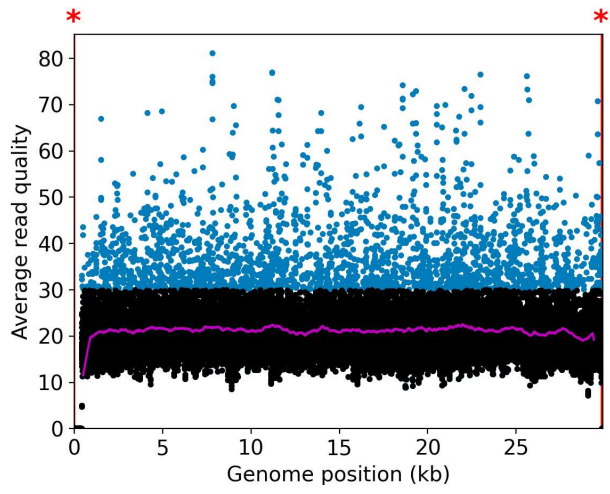
[CFSANSMP000119229](#)



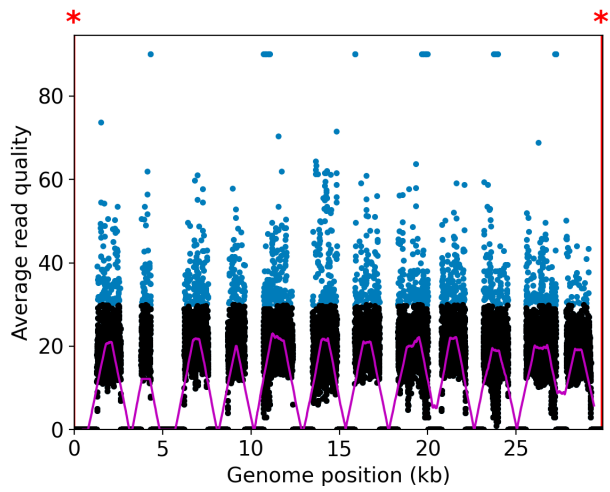
[CFSANSMP000119231](#)



[unclassified](#)



Water



**CFSAN/OAO
BIostatistics AND BIOinformatics STAFF**

WASTEWATER SARS-COV2 ANALYSIS REPORT

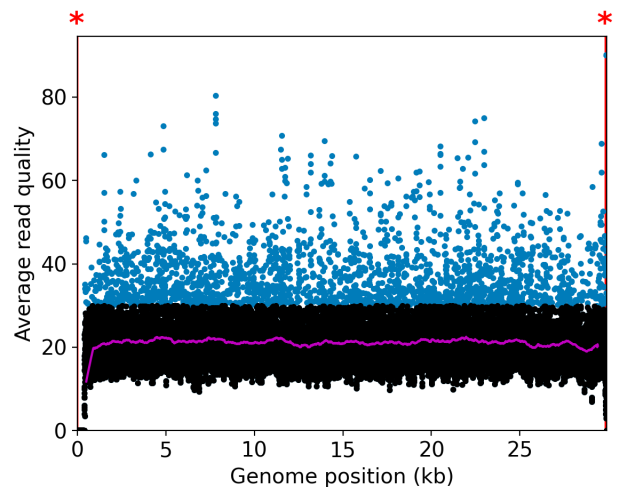
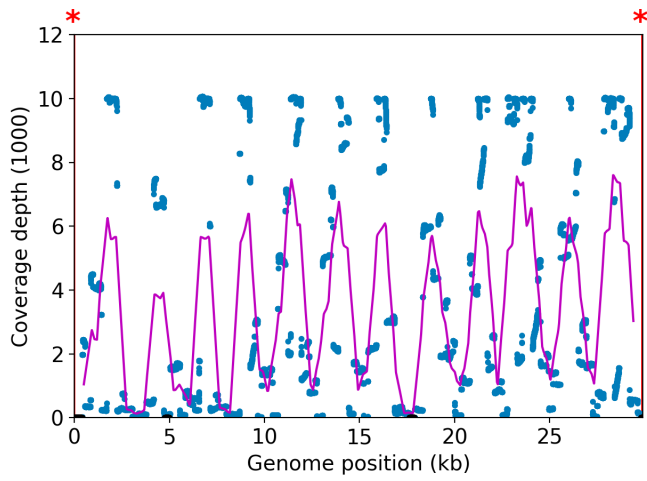
Sample name:	CFSANSMP000119220
Date generated:	2023-02-02, 09:14:16 EST
Timestamp of C-WAP version used:	Tue Jan 31 11:49:22 2023 -0500
Executed by:	Jasmine Amirzadegan (Jasmine.Amirzadegan@fda.hhs.gov)
Executed on:	172.20.44.125 (aka n125.raven.cfsan)

Sequencing summary

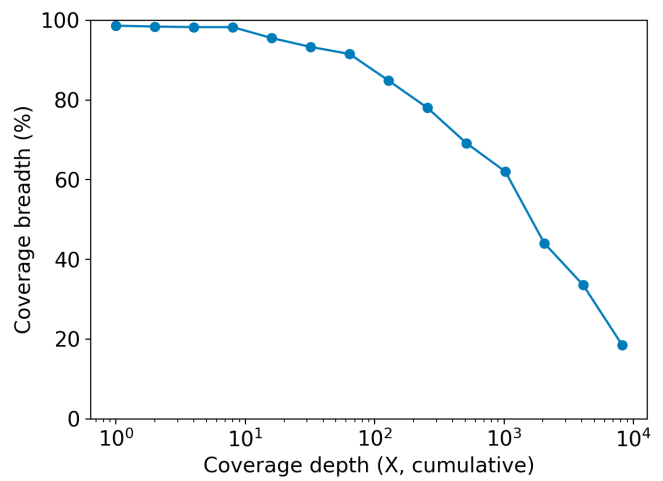
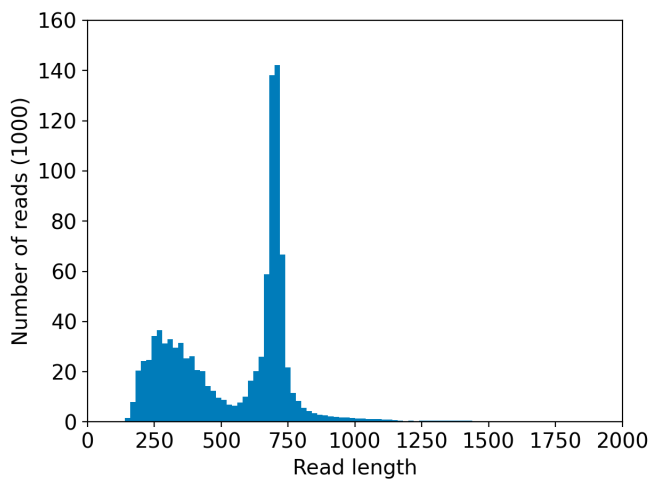
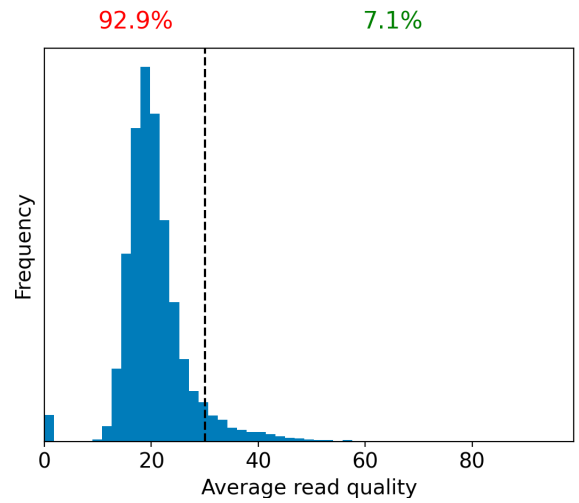
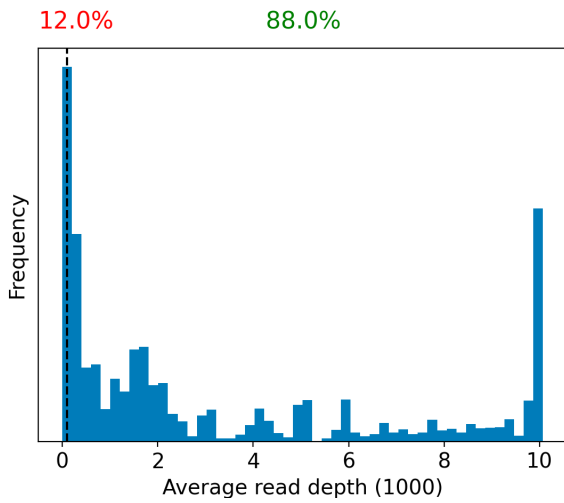
Sequencing chemistry:	Missing with Missing
Source site:	Missing (?.?)
Sampling date:	Missing
Collected by:	Missing
Sequenced by:	Missing
Total number of reads:	1240436
Reads aligned:	748711 (60%)
Average read quality:	19.8
Average read length:	693
Reads passing filter:	748699 (60%)
Average read quality passing filter:	19.8
Average read length passing filter:	693
Average coverage passing filter:	17351X

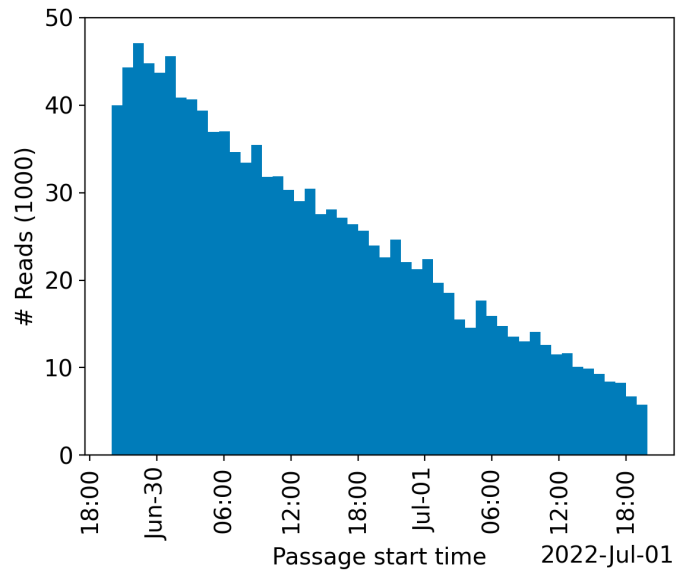
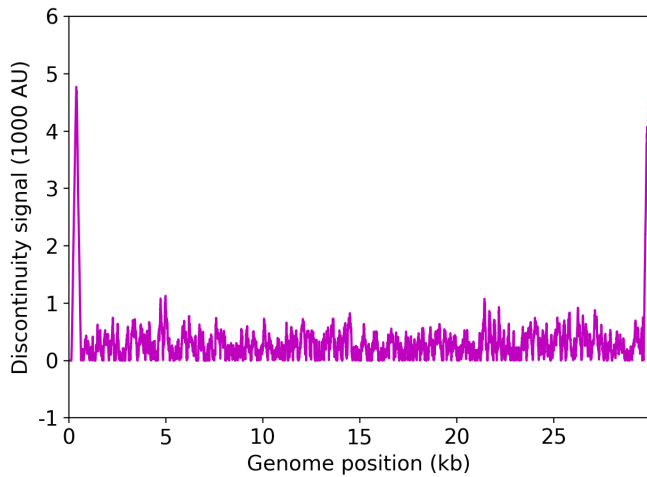
A read passes filter if the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4.

Overall sequence characteristics



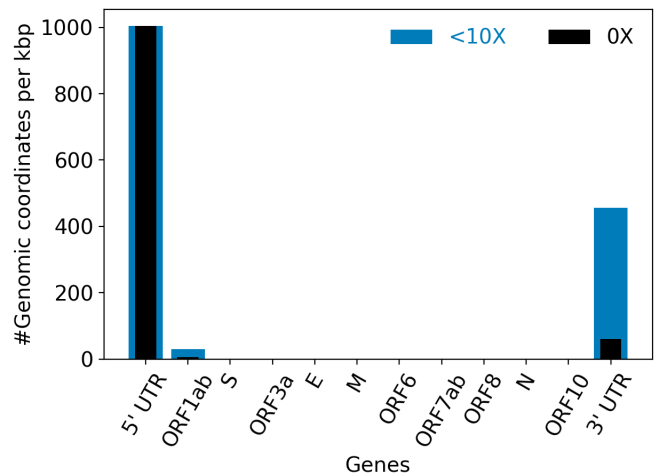
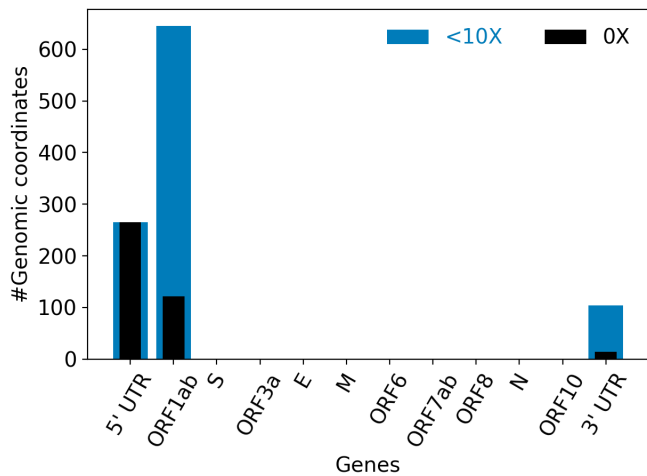
NOTE: The red shaded areas marked with a (*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.





	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	152nt (0%)	152nt (0%)
All genomic coordinates:	401nt (1%)	1014nt (3%)
Common SNPs:	1nt (2%)	1nt (2%)
Diverse SNPs:	2nt (0%)	29nt (12%)
Rare SNPs:	49nt (5%)	66nt (7%)

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.



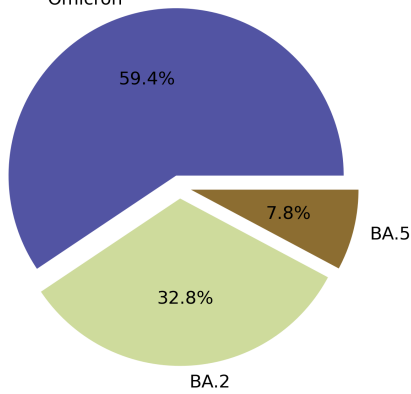
Hits to SARS-Cov2 genome (kraken2):	743646 reads (59.95%)
Hits to human genome (kraken2):	18681 reads (1.51%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
	Coronaviridae (59.95%)

Most abundant organisms (kraken2, family level):

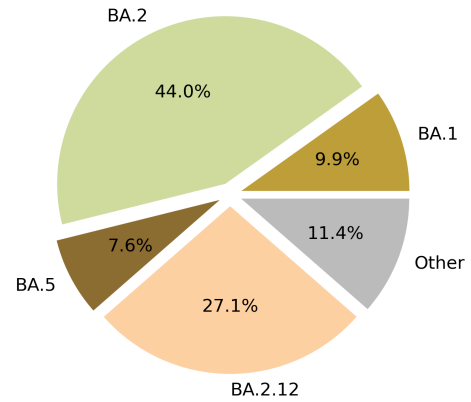
Hominidae (1.51%)
Arcobacteraceae (0.85%)
Enterobacteriaceae (0.37%)
Comamonadaceae (0.25%)

Detected variants (Experimental)

Abundance of variants by linear regression



Abundance of variants by kallisto

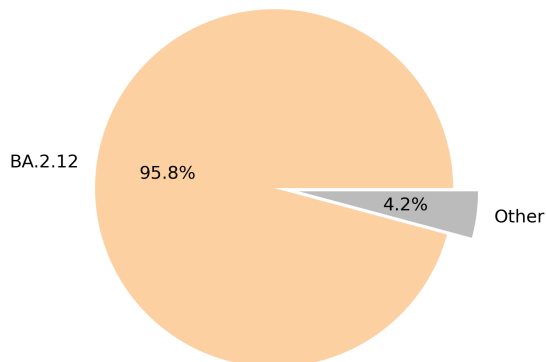


Based on deconvolution, [B.1.1.529](#) is estimated to constitute 59.39% of the viral particles and hence is the most abundant variant in the sample. The R^2 for the linear regression was 0.63. Variants that were detected less than 5% were grouped under "Other"

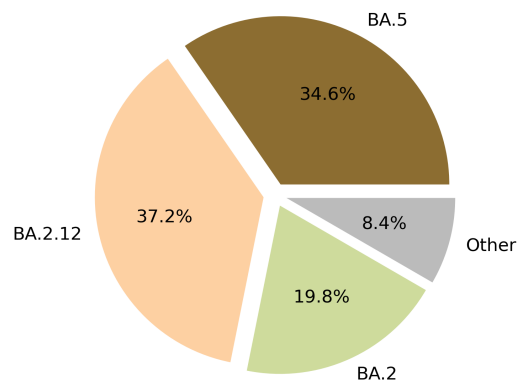
Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [BA.2](#) lineage. If this is a sample consisting of a single source of pathogens or an overwhelming majority of the different sources are infected with the same variant, the sample is dominated by this variant.

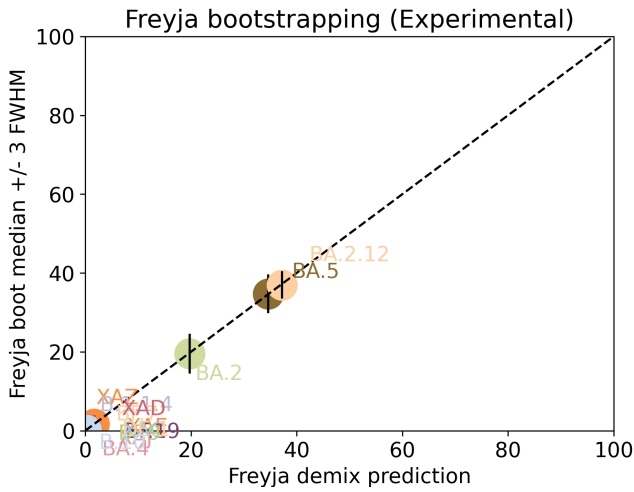
Based on mapping individual reads to the variant consensus sequences in the reference database, kallisto predicts that the sample is dominated by [BA.2](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.

Abundance of variants by kraken2+bracken, using majorCovid DB



Abundance of variants by Freyja





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	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
Characteristic mutations detected	(3 of 13) S:G142D S:L452R S:T478K	(5 of 26) M:D3G NUC:C25000T NUC:C25584T ORF1AB:I3758V S:Q493R	(23 of 31) N:S413R NUC:A20055G NUC:A9424G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:L3027F ORF1AB:L3201F ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:Q493R S:R408S S:S371F S:T19I S:T376A	(11 of 21) N:S413R NUC:C12880T NUC:C15714T NUC:C26858T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I S:D405N S:Q493R S:S371F	(22 of 31) N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A NUC:G27788T ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G	(20 of 28) N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G

[Jaccard Index](#) is a measure of similarity between two sets A and B, reaching the maximum value of 1 if A=B and minimum value of 0 if $A \cap B = \{\}$. In the c(d) representation below, c represents the Jaccard index of the set of mutations that were experimentally detected for this sample as listed above, whereas d refers to the ideal value of the Jaccard index expected from complete genome coverage without any sequencing errors.

	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
B.1.617.2	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.04 (0.02)	0.05 (0.03)
BA.1	0.00 (0.00)	1.00 (1.00)	0.12 (0.10)	0.07 (0.21)	0.08 (0.08)	0.09 (0.08)
BA.2	0.00 (0.00)	0.12 (0.10)	1.00 (1.00)	0.48 (0.33)	0.67 (0.63)	0.65 (0.59)
BA.3	0.00 (0.00)	0.07 (0.21)	0.48 (0.33)	1.00 (1.00)	0.43 (0.30)	0.41 (0.29)
BA.4	0.04 (0.02)	0.08 (0.08)	0.67 (0.63)	0.43 (0.30)	1.00 (1.00)	0.91 (0.84)

BA.5

0.05 ([0.03](#))

0.09 ([0.08](#))

0.65 ([0.59](#))

0.41 ([0.29](#))

0.91 ([0.84](#))

1.00 ([1.00](#))

Detected mutations

Excluded from this pdf version due to file size limitations.

CFSAN/OAO
BIostatistics AND Bioinformatics Staff

WASTEWATER SARS-COV2 ANALYSIS REPORT

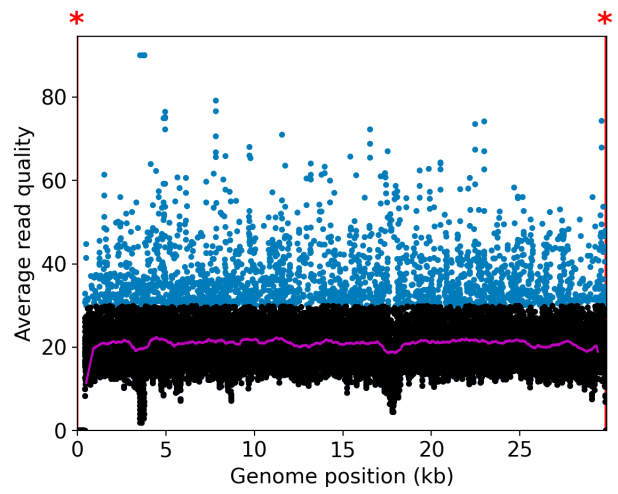
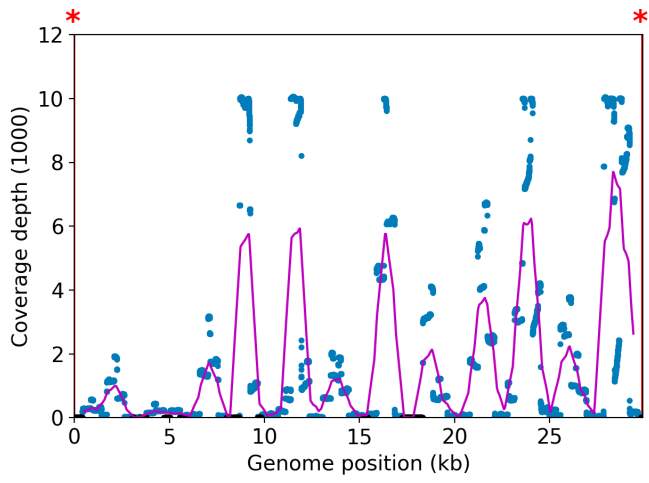
Sample name:	CFSANSMP000119221
Date generated:	2023-02-02, 09:00:28 EST
Timestamp of C-WAP version used:	Tue Jan 31 11:49:22 2023 -0500
Executed by:	Jasmine Amirzadegan (Jasmine.Amirzadegan@fda.hhs.gov)
Executed on:	172.20.44.121 (aka n121.raven.cfsan)

Sequencing summary

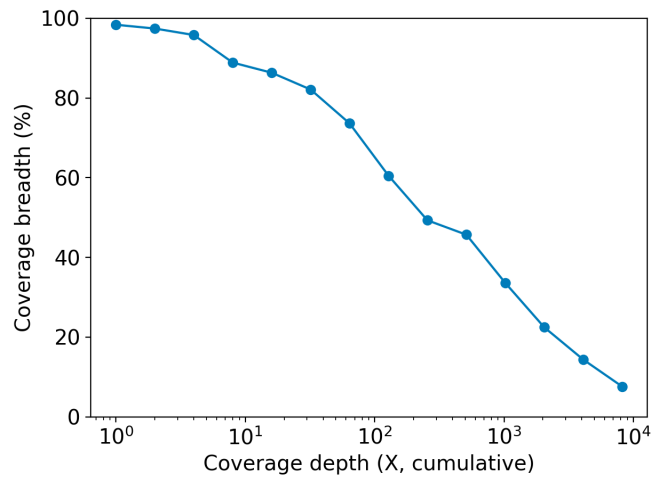
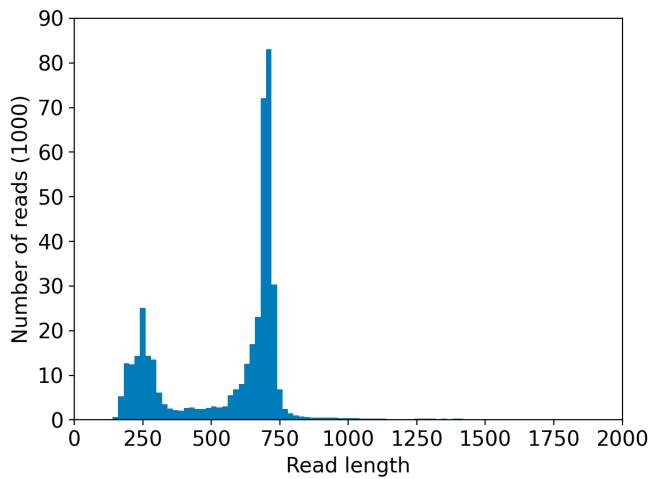
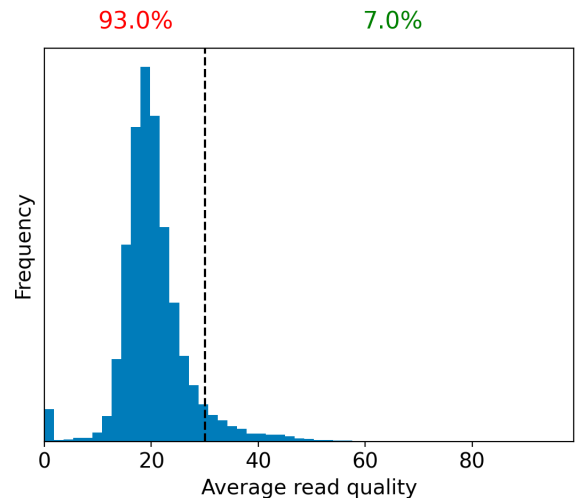
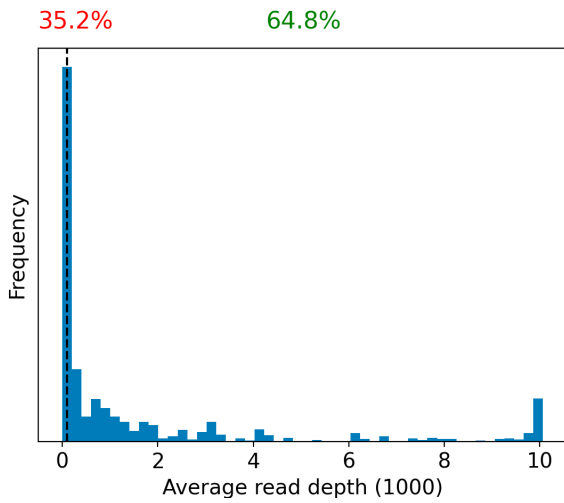
Sequencing chemistry:	Missing with Missing
Source site:	Missing (?.?)
Sampling date:	Missing
Collected by:	Missing
Sequenced by:	Missing
Total number of reads:	420127
Reads aligned:	304558 (72%)
Average read quality:	19.3
Average read length:	665
Reads passing filter:	304556 (72%)
Average read quality passing filter:	19.3
Average read length passing filter:	665
Average coverage passing filter:	6772X

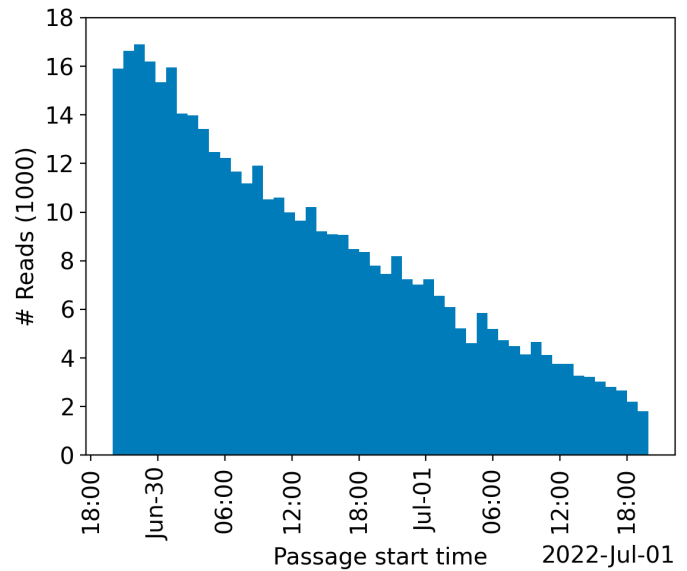
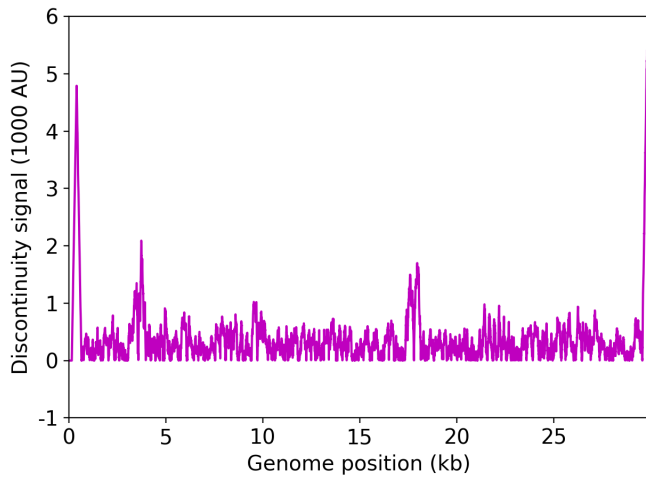
A read passes filter if the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4.

Overall sequence characteristics



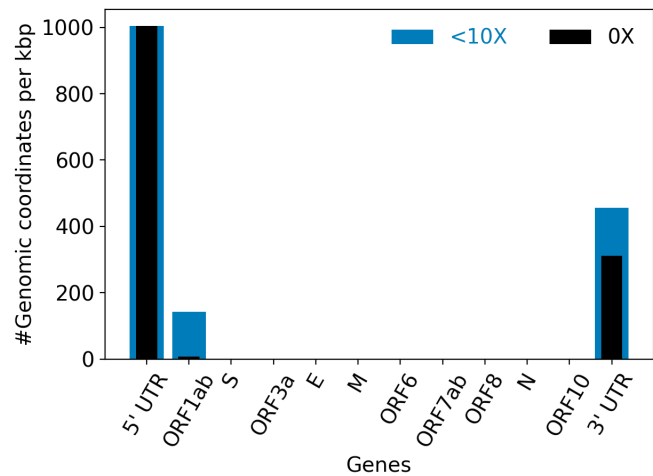
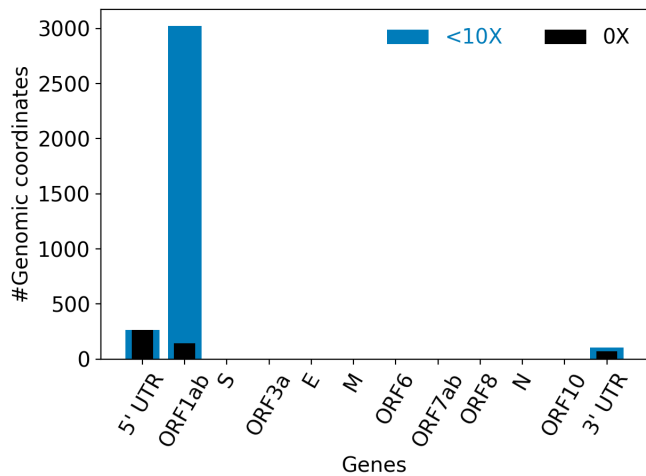
NOTE: The red shaded areas marked with a (*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.





	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	152nt (0%)	152nt (0%)
All genomic coordinates:	481nt (1%)	3391nt (11%)
Common SNPs:	1nt (2%)	2nt (5%)
Diverse SNPs:	29nt (12%)	32nt (14%)
Rare SNPs:	61nt (6%)	66nt (7%)

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.

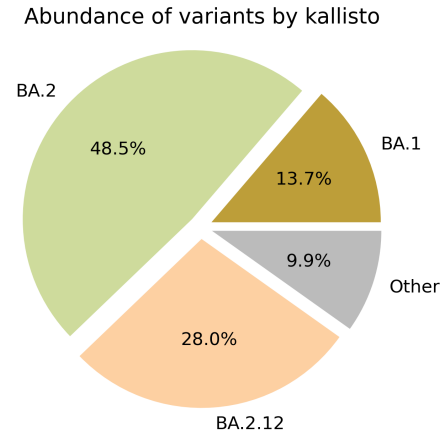
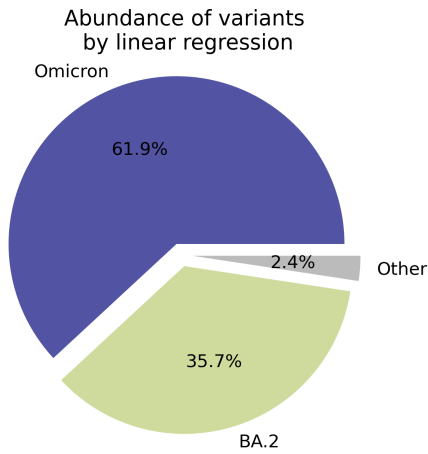


Hits to SARS-Cov2 genome (kraken2):	301382 reads (71.74%)
Hits to human genome (kraken2):	4549 reads (1.08%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
	Coronaviridae (71.74%)

Most abundant organisms (kraken2, family level):

Arcobacteraceae (1.84%)
Hominidae (1.08%)
Enterobacteriaceae (0.43%)
Bacteroidaceae (0.36%)

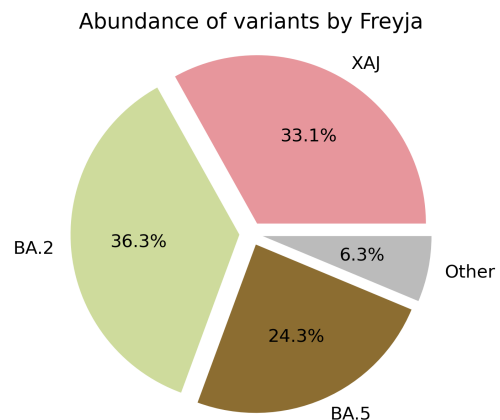
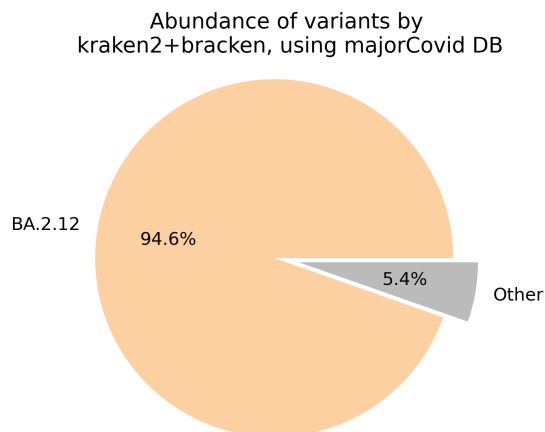
Detected variants (Experimental)

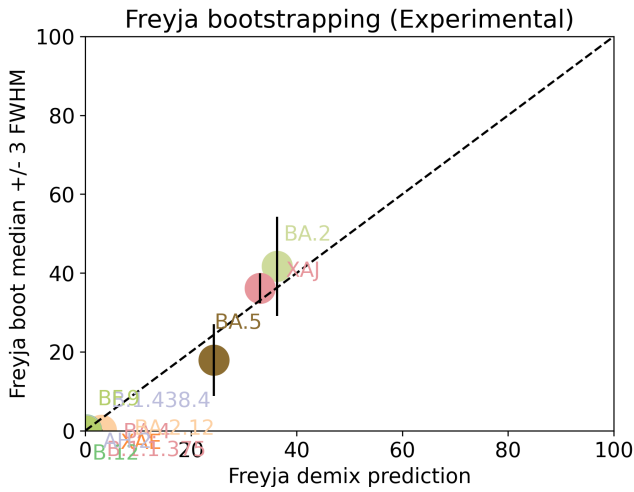


Based on deconvolution, [B.1.1.529](#) is estimated to constitute 61.94% of the viral particles and hence is the most abundant variant in the sample. The R^2 for the linear regression was 0.64. Variants that were detected less than 5% were grouped under "Other"

Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [BA.2](#) 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.2](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.





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	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
Characteristic mutations detected	(3 of 13) S:G142D S:L452R S:T478K	(5 of 26) M:D3G NUC:C25000T NUC:C25584T ORF1AB:I3758V S:Q493R	(22 of 31) N:S413R NUC:A20055G NUC:A9424G NUC:C10198T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:L3027F ORF1AB:L3201F ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:Q493R S:R408S S:S371F S:T19I S:T376A	(10 of 21) N:S413R NUC:C15714T NUC:C26858T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I S:D405N S:Q493R S:S371F	(20 of 31) N:S413R NUC:A20055G NUC:C10198T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C4321T NUC:G10447A NUC:G27788T ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A	(19 of 28) M:D3N N:S413R NUC:A20055G NUC:C10198T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A

Jaccard Index is a measure of similarity between two sets A and B, reaching the maximum value of 1 if A=B and minimum value of 0 if $A \cap B = \{\}$. In the c(d) representation below, c represents the Jaccard index of the set of mutations that were experimentally detected for this sample as listed above, whereas d refers to the ideal value of the Jaccard index expected from complete genome coverage without any sequencing errors.

	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
B.1.617.2	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.05 (0.02)	0.05 (0.03)
BA.1	0.00 (0.00)	1.00 (1.00)	0.12 (0.10)	0.07 (0.21)	0.09 (0.08)	0.09 (0.08)
BA.2	0.00 (0.00)	0.12 (0.10)	1.00 (1.00)	0.45 (0.33)	0.68 (0.63)	0.64 (0.59)
BA.3	0.00 (0.00)	0.07 (0.21)	0.45 (0.33)	1.00 (1.00)	0.43 (0.30)	0.38 (0.29)
BA.4	0.05 (0.02)	0.09 (0.08)	0.68 (0.63)	0.43 (0.30)	1.00 (1.00)	0.86 (0.84)

BA.5

0.05 ([0.03](#))

0.09 ([0.08](#))

0.64 ([0.59](#))

0.38 ([0.29](#))

0.86 ([0.84](#))

1.00 ([1.00](#))

Detected mutations

Excluded from this pdf version due to file size limitations.

**CFSAN/OAO
BIostatistics AND BIOinformatics STAFF**

WASTEWATER SARS-COV2 ANALYSIS REPORT

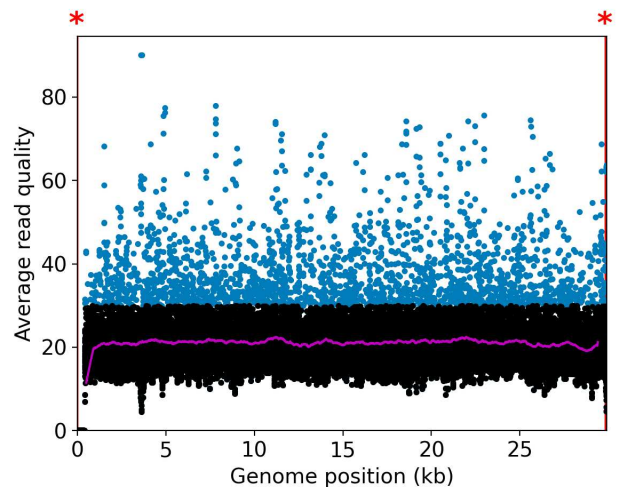
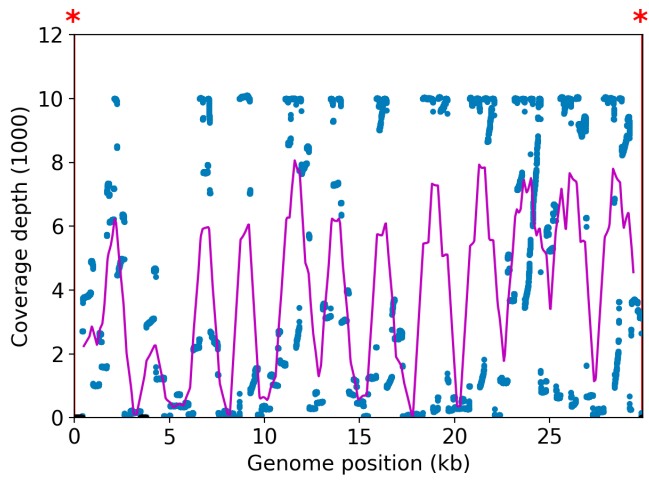
Sample name:	CFSANSMP000119222
Date generated:	2023-02-02, 09:12:45 EST
Timestamp of C-WAP version used:	Tue Jan 31 11:49:22 2023 -0500
Executed by:	Jasmine Amirzadegan (Jasmine.Amirzadegan@fda.hhs.gov)
Executed on:	172.20.44.125 (aka n125.raven.cfsan)

Sequencing summary

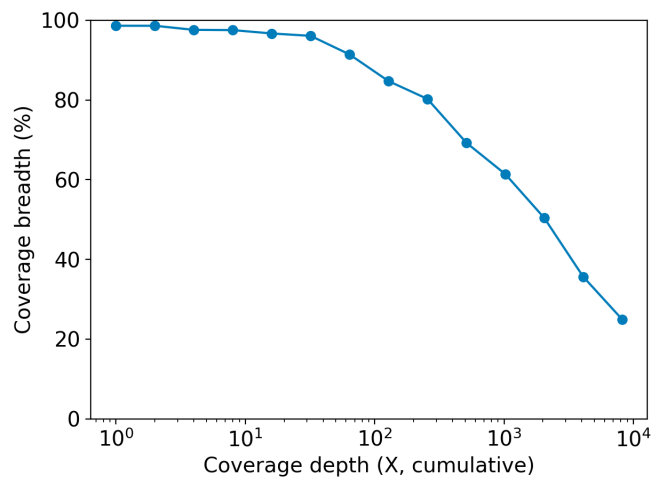
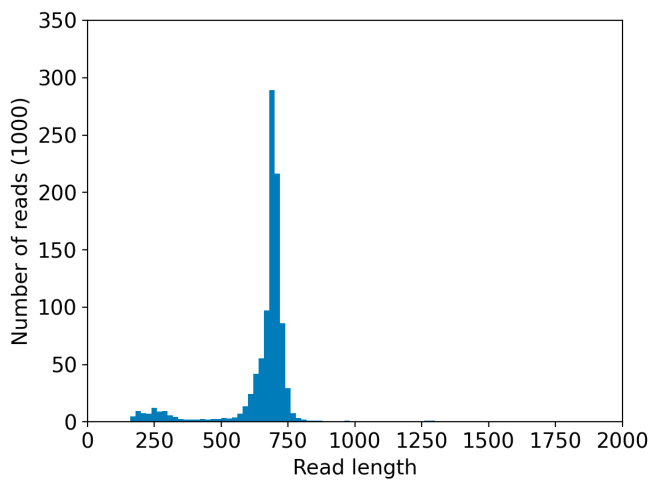
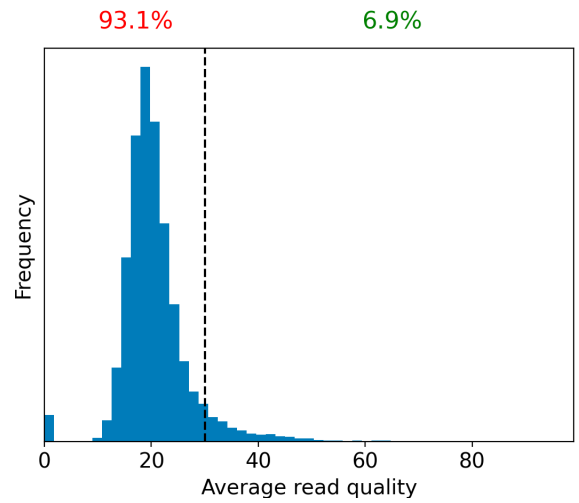
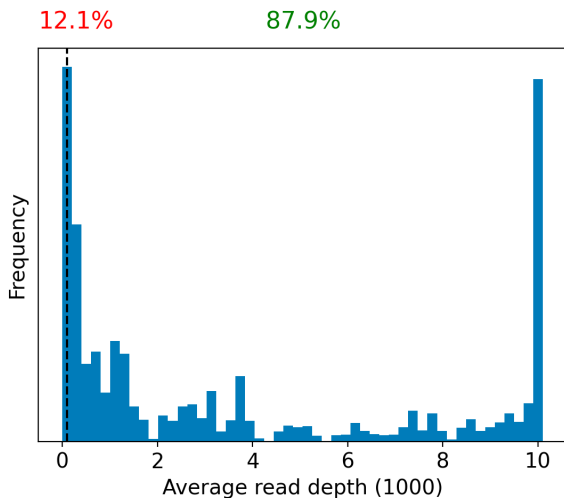
Sequencing chemistry:	Missing with Missing
Source site:	Missing (?.?)
Sampling date:	Missing
Collected by:	Missing
Sequenced by:	Missing
Total number of reads:	1536416
Reads aligned:	1447547 (94%)
Average read quality:	19.8
Average read length:	684
Reads passing filter:	1447541 (94%)
Average read quality passing filter:	19.8
Average read length passing filter:	684
Average coverage passing filter:	33110X

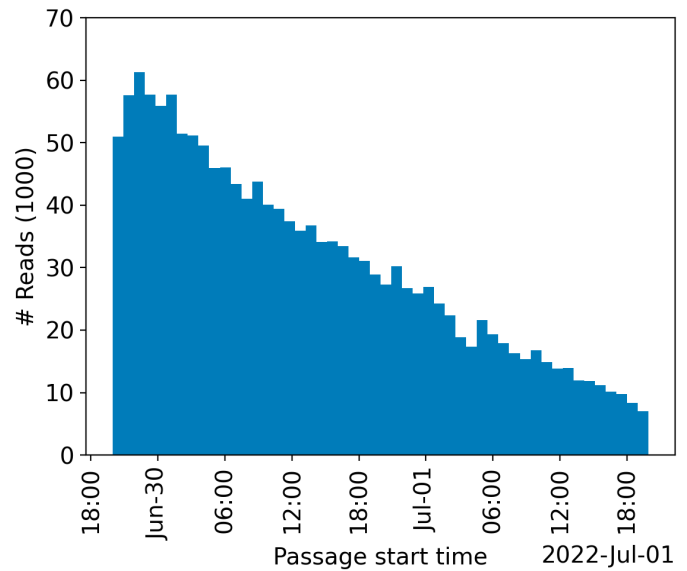
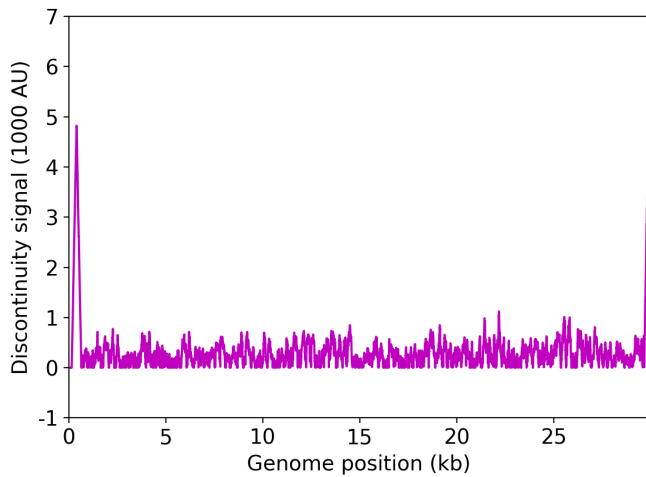
A read passes filter if the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4.

Overall sequence characteristics



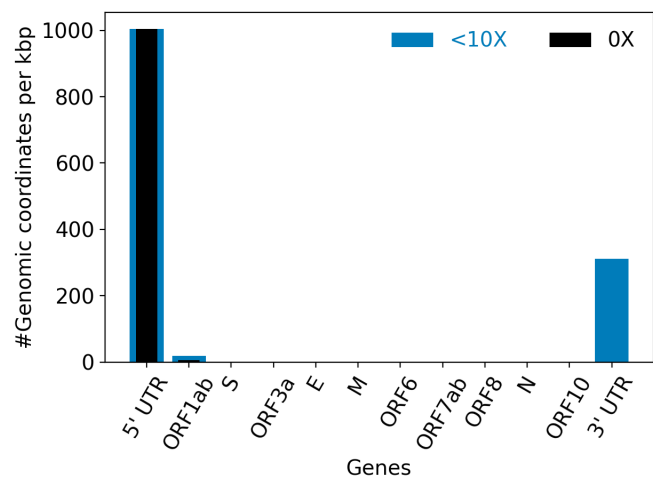
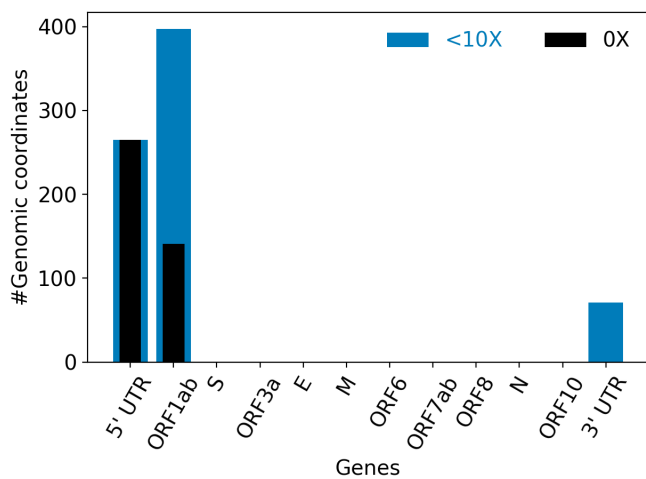
NOTE: The red shaded areas marked with a (*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.





	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	152nt (0%)	152nt (0%)
All genomic coordinates:	406nt (1%)	733nt (2%)
Common SNPs:	1nt (2%)	1nt (2%)
Diverse SNPs:	2nt (0%)	29nt (12%)
Rare SNPs:	49nt (5%)	61nt (6%)

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.



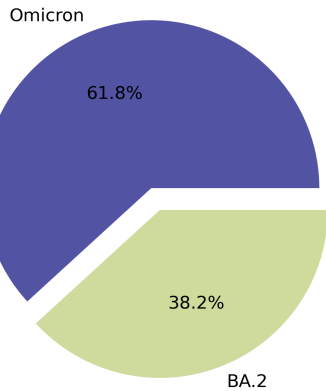
Hits to SARS-Cov2 genome (kraken2):	1435289 reads (93.42%)
Hits to human genome (kraken2):	7983 reads (0.52%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
	Coronaviridae (93.42%)

Most abundant organisms (kraken2, family level):

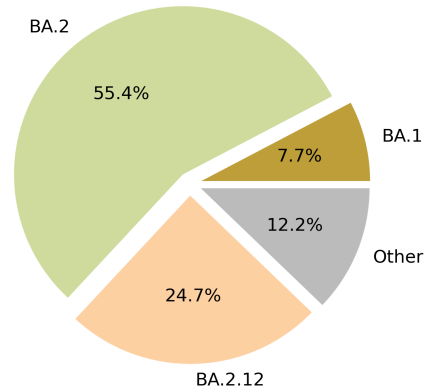
Hominidae (0.52%)
Arcobacteraceae (0.48%)
Enterobacteriaceae (0.10%)
Bacteroidaceae (0.06%)

Detected variants (Experimental)

Abundance of variants by linear regression



Abundance of variants by kallisto

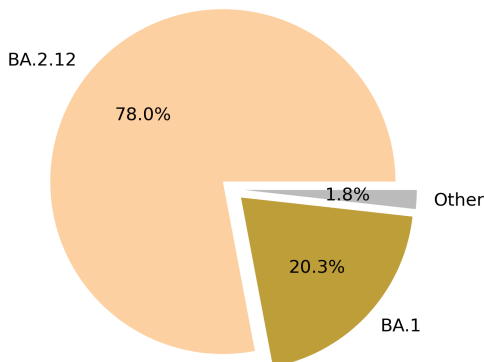


Based on deconvolution, [B.1.1.529](#) is estimated to constitute 61.81% of the viral particles and hence is the most abundant variant in the sample. The R^2 for the linear regression was 0.64. Variants that were detected less than 5% were grouped under "Other"

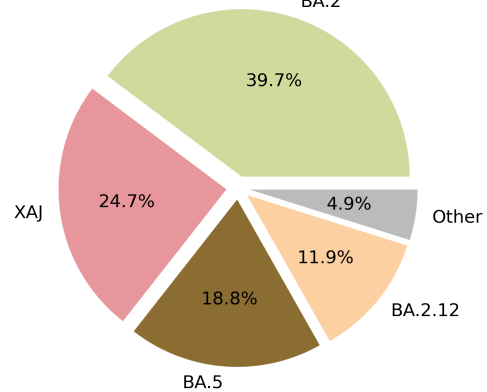
Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [BA.2](#) lineage. If this is a sample consisting of a single source of pathogens or an overwhelming majority of the different sources are infected with the same variant, the sample is dominated by this variant.

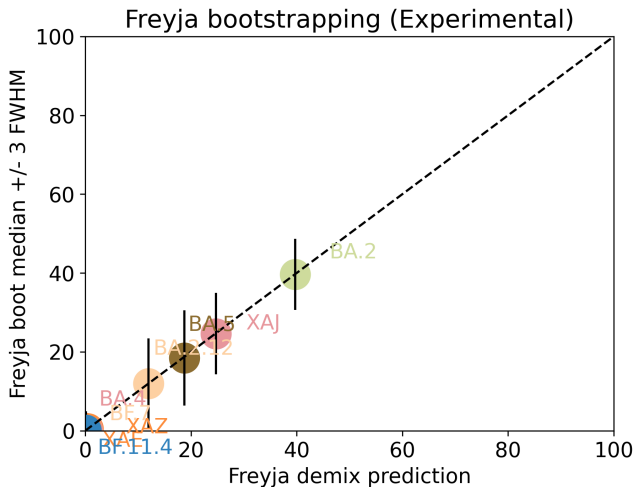
Based on mapping individual reads to the variant consensus sequences in the reference database, kallisto predicts that the sample is dominated by [BA.2](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.

Abundance of variants by kraken2+bracken, using majorCovid DB



Abundance of variants by Freyja





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	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
Characteristic mutations detected	(3 of 13) S:G142D S:L452R S:T478K	(5 of 26) M:D3G NUC:C25000T NUC:C25584T ORF1AB:I3758V S:Q493R	(22 of 31) N:S413R NUC:A9424G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:L3027F ORF1AB:L3201F ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:Q493R S:R408S S:S371F S:T19I S:T376A	(11 of 21) N:S413R NUC:C12880T NUC:C15714T NUC:C26858T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I S:D405N S:Q493R S:S371F	(20 of 31) N:S413R NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A NUC:G27788T ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A	(18 of 28) N:S413R NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A

Jaccard Index is a measure of similarity between two sets A and B, reaching the maximum value of 1 if A=B and minimum value of 0 if $A \cap B = \{\}$. In the c(d) representation below, c represents the Jaccard index of the set of mutations that were experimentally detected for this sample as listed above, whereas d refers to the ideal value of the Jaccard index expected from complete genome coverage without any sequencing errors.

	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
B.1.617.2	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.05 (0.02)	0.05 (0.03)
BA.1	0.00 (0.00)	1.00 (1.00)	0.12 (0.10)	0.07 (0.21)	0.09 (0.08)	0.10 (0.08)
BA.2	0.00 (0.00)	0.12 (0.10)	1.00 (1.00)	0.50 (0.33)	0.68 (0.63)	0.67 (0.59)
BA.3	0.00 (0.00)	0.07 (0.21)	0.50 (0.33)	1.00 (1.00)	0.48 (0.30)	0.45 (0.29)
BA.4	0.05 (0.02)	0.09 (0.08)	0.68 (0.63)	0.48 (0.30)	1.00 (1.00)	0.90 (0.84)

BA.5

0.05 ([0.03](#))

0.10 ([0.08](#))

0.67 ([0.59](#))

0.45 ([0.29](#))

0.90 ([0.84](#))

1.00 ([1.00](#))

Detected mutations

Excluded from this pdf version due to file size limitations.

CFSAN/OAO
BIostatistics AND Bioinformatics Staff

WASTEWATER SARS-COV2 ANALYSIS REPORT

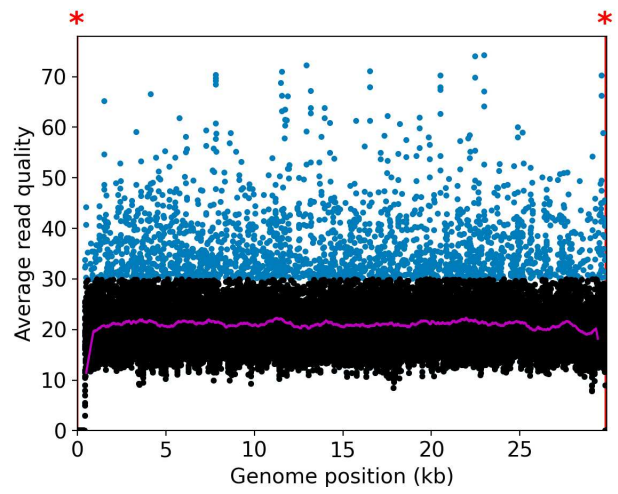
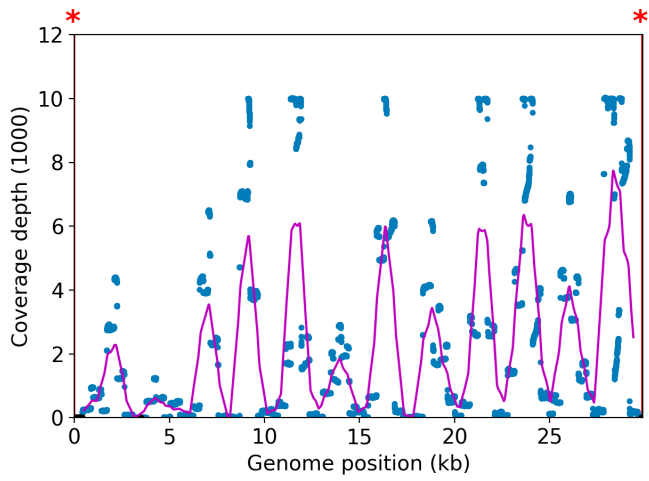
Sample name:	CFSANSMP000119223
Date generated:	2023-02-02, 09:14:44 EST
Timestamp of C-WAP version used:	Tue Jan 31 11:49:22 2023 -0500
Executed by:	Jasmine Amirzadegan (Jasmine.Amirzadegan@fda.hhs.gov)
Executed on:	172.20.44.125 (aka n125.raven.cfsan)

Sequencing summary

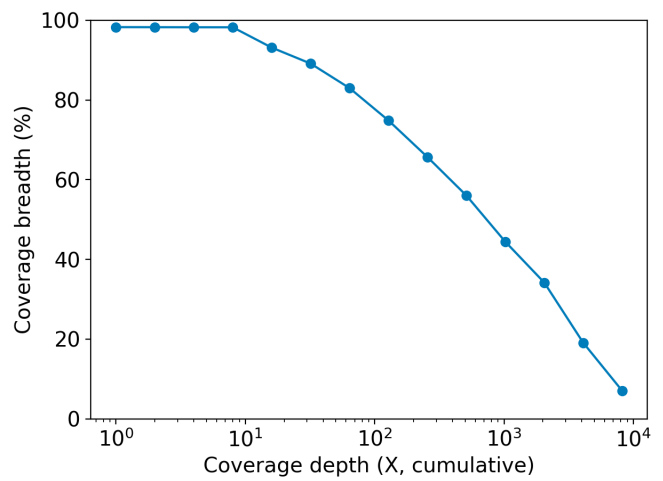
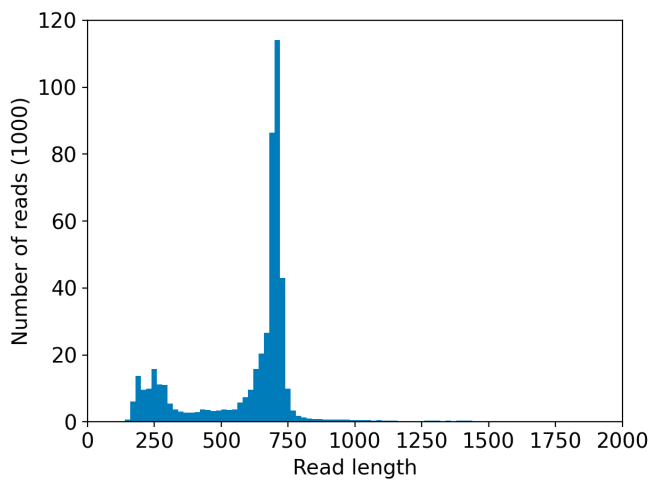
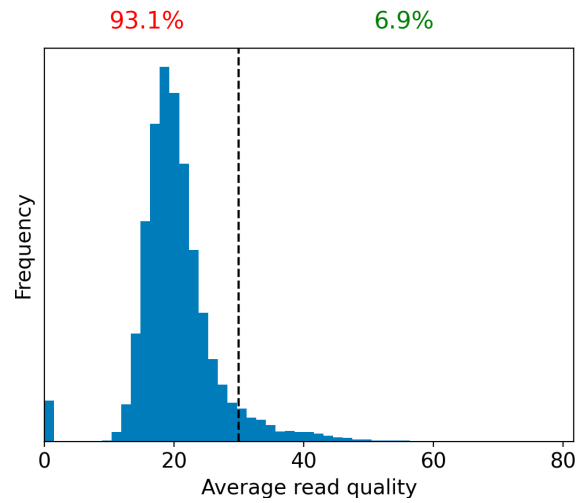
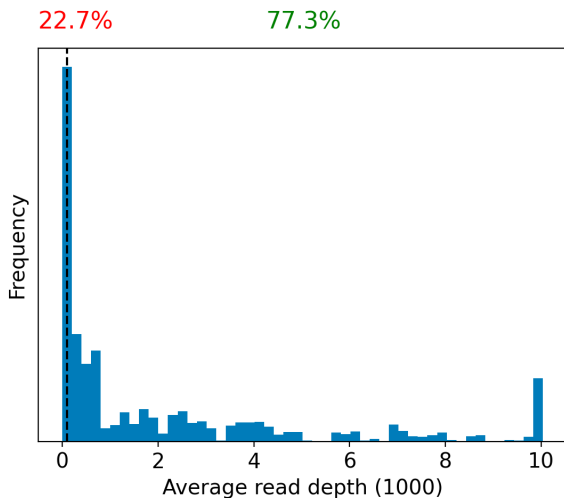
Sequencing chemistry:	Missing with Missing
Source site:	Missing (?.?)
Sampling date:	Missing
Collected by:	Missing
Sequenced by:	Missing
Total number of reads:	487373
Reads aligned:	387337 (79%)
Average read quality:	19.8
Average read length:	668
Reads passing filter:	387336 (79%)
Average read quality passing filter:	19.8
Average read length passing filter:	668
Average coverage passing filter:	8652X

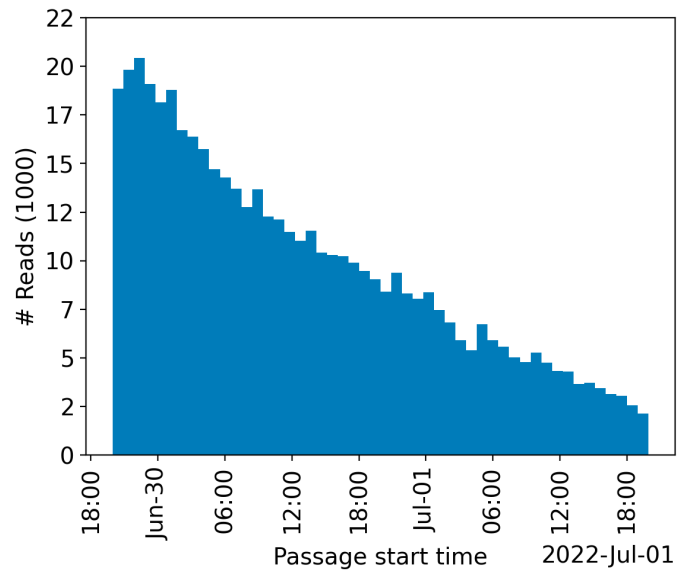
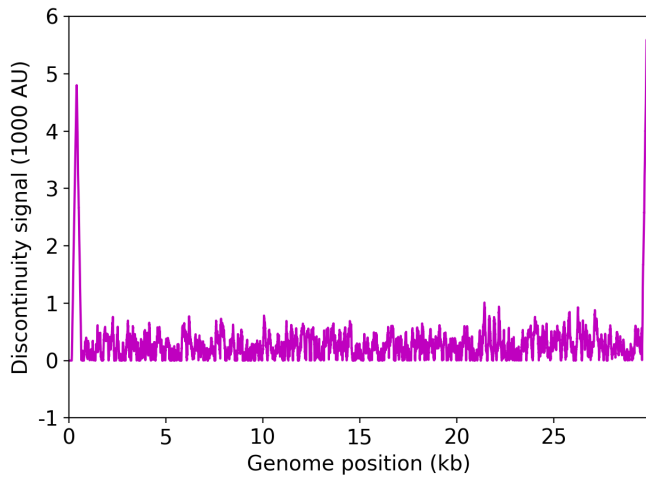
A read passes filter if the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4.

Overall sequence characteristics



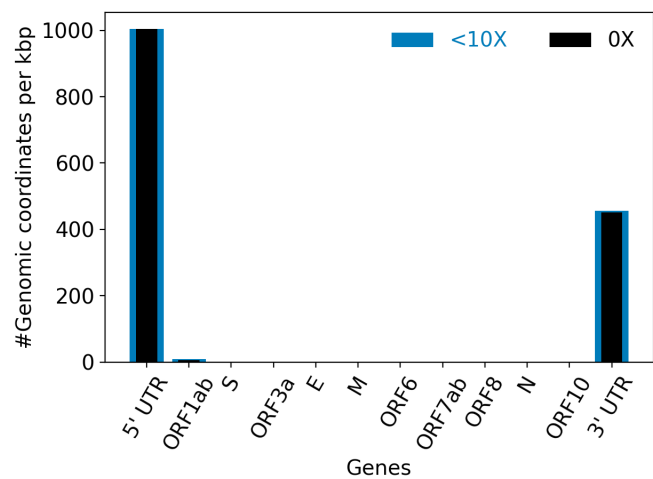
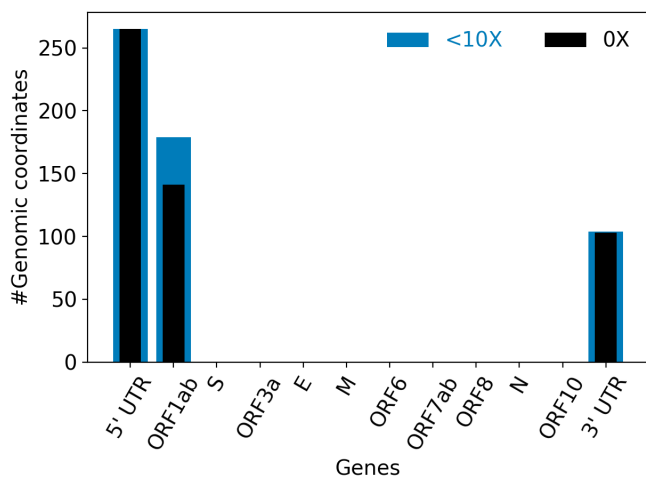
NOTE: The red shaded areas marked with a (*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.





	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	152nt (0%)	152nt (0%)
All genomic coordinates:	509nt (1%)	548nt (1%)
Common SNPs:	1nt (2%)	1nt (2%)
Diverse SNPs:	29nt (12%)	29nt (12%)
Rare SNPs:	66nt (7%)	66nt (7%)

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.



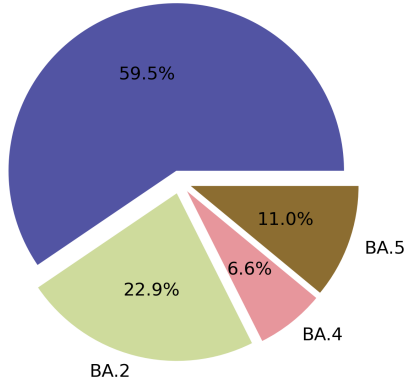
Hits to SARS-Cov2 genome (kraken2):	382881 reads (78.56%)
Hits to human genome (kraken2):	4425 reads (0.91%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
	Coronaviridae (78.56%)

Most abundant organisms (kraken2, family level):

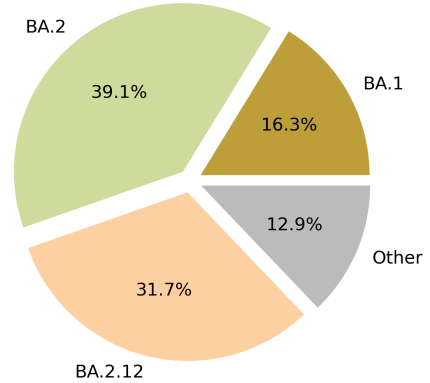
Arcobacteraceae (3.00%)
Hominidae (0.91%)
Bacteroidaceae (0.44%)
Enterobacteriaceae (0.40%)

Detected variants (Experimental)

Abundance of variants
by linear regression
Omicron



Abundance of variants by kallisto

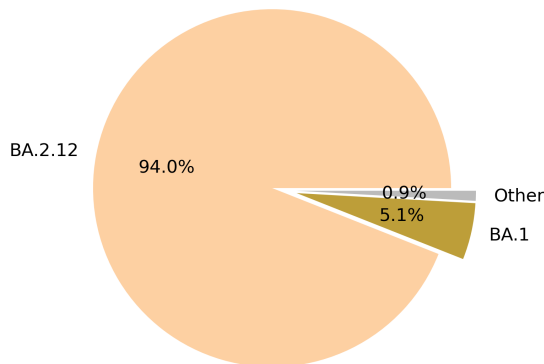


Based on deconvolution, [B.1.1.529](#) is estimated to constitute 59.47% of the viral particles and hence is the most abundant variant in the sample. The R^2 for the linear regression was 0.62. Variants that were detected less than 5% were grouped under "Other"

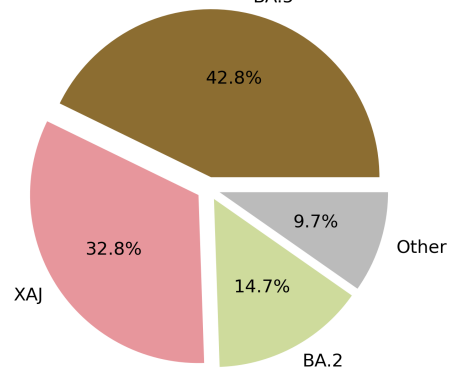
Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [BA.2](#) lineage. If this is a sample consisting of a single source of pathogens or an overwhelming majority of the different sources are infected with the same variant, the sample is dominated by this variant.

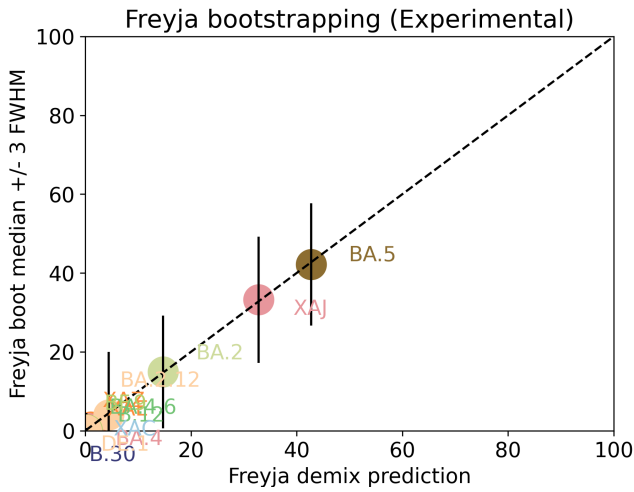
Based on mapping individual reads to the variant consensus sequences in the reference database, kallisto predicts that the sample is dominated by [BA.2](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.

Abundance of variants by
kraken2+bracken, using majorCovid DB



Abundance of variants by Freyja
BA.5





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	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
Characteristic mutations detected	(3 of 13) S:G142D S:L452R S:T478K	(5 of 26) M:D3G NUC:C25000T NUC:C25584T ORF1AB:I3758V S:Q493R	(23 of 31) N:S413R NUC:A20055G NUC:A9424G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:L3027F ORF1AB:L3201F ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:Q493R S:R408S S:S371F S:T19I S:T376A	(11 of 21) N:S413R NUC:C12880T NUC:C15714T NUC:C26858T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I S:D405N S:Q493R S:S371F	(22 of 31) N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A NUC:G27788T ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G	(21 of 28) M:D3N N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G

[Jaccard Index](#) is a measure of similarity between two sets A and B, reaching the maximum value of 1 if A=B and minimum value of 0 if $A \cap B = \{\}$. In the c(d) representation below, c represents the Jaccard index of the set of mutations that were experimentally detected for this sample as listed above, whereas d refers to the ideal value of the Jaccard index expected from complete genome coverage without any sequencing errors.

	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
B.1.617.2	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.04 (0.02)	0.04 (0.03)
BA.1	0.00 (0.00)	1.00 (1.00)	0.12 (0.10)	0.07 (0.21)	0.08 (0.08)	0.08 (0.08)
BA.2	0.00 (0.00)	0.12 (0.10)	1.00 (1.00)	0.48 (0.33)	0.67 (0.63)	0.63 (0.59)
BA.3	0.00 (0.00)	0.07 (0.21)	0.48 (0.33)	1.00 (1.00)	0.43 (0.30)	0.39 (0.29)
BA.4	0.04 (0.02)	0.08 (0.08)	0.67 (0.63)	0.43 (0.30)	1.00 (1.00)	0.87 (0.84)

BA.5

0.04 ([0.03](#))

0.08 ([0.08](#))

0.63 ([0.59](#))

0.39 ([0.29](#))

0.87 ([0.84](#))

1.00 ([1.00](#))

Detected mutations

Excluded from this pdf version due to file size limitations.

CFSAN/OAO
BIostatistics AND BIOinformatics STAFF

WASTEWATER SARS-COV2 ANALYSIS REPORT

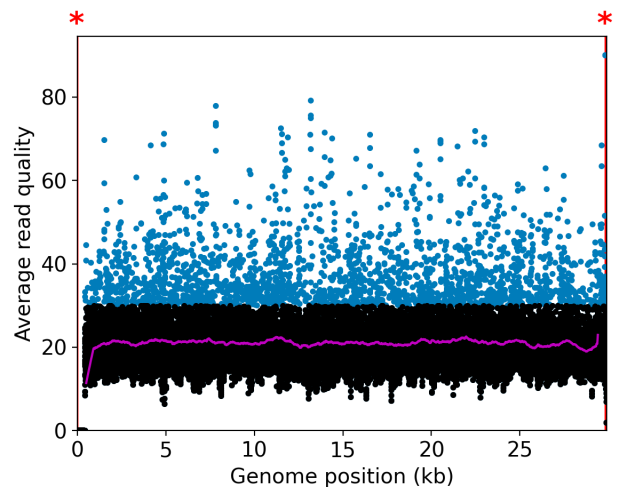
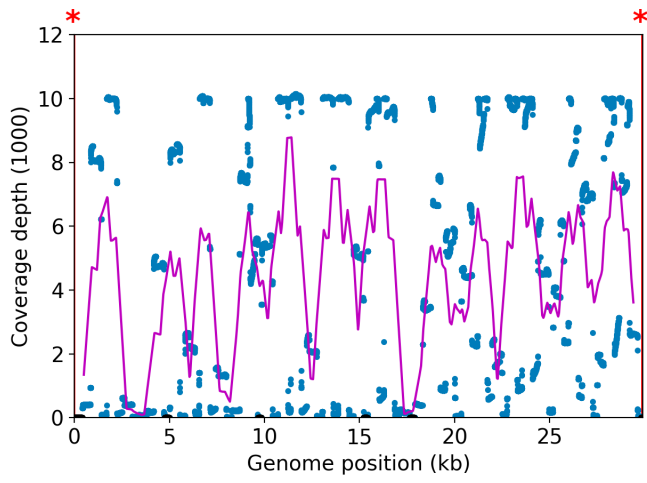
Sample name:	CFSANSMP000119224
Date generated:	2023-02-02, 09:09:31 EST
Timestamp of C-WAP version used:	Tue Jan 31 11:49:22 2023 -0500
Executed by:	Jasmine Amirzadegan (Jasmine.Amirzadegan@fda.hhs.gov)
Executed on:	172.20.44.125 (aka n125.raven.cfsan)

Sequencing summary

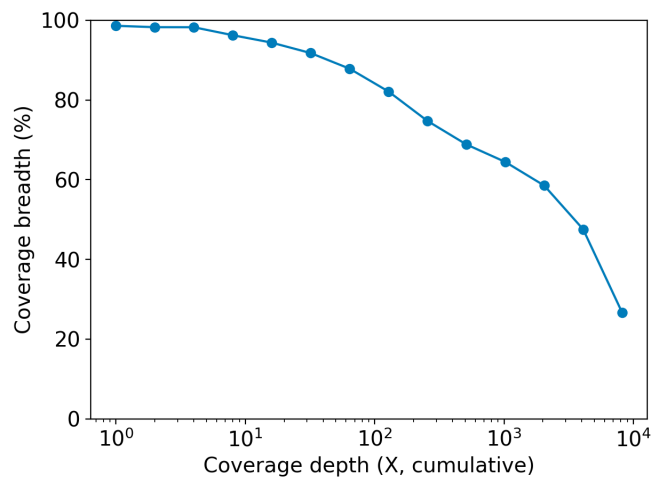
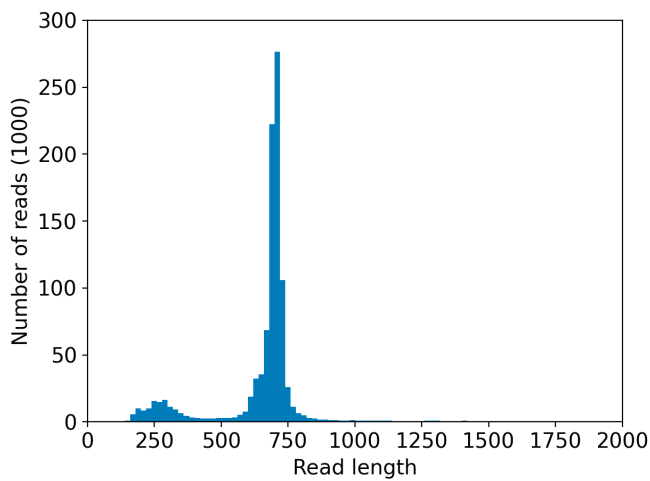
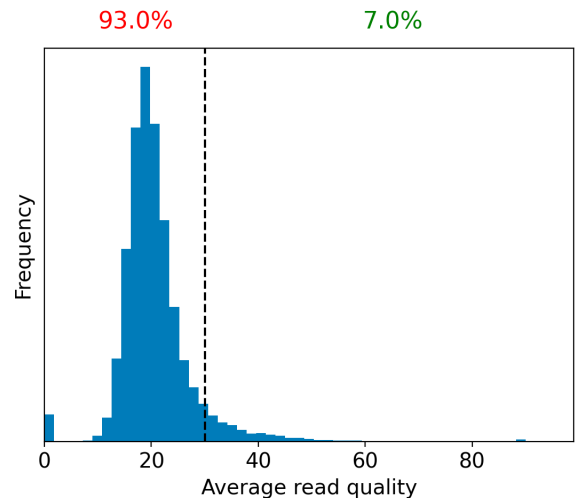
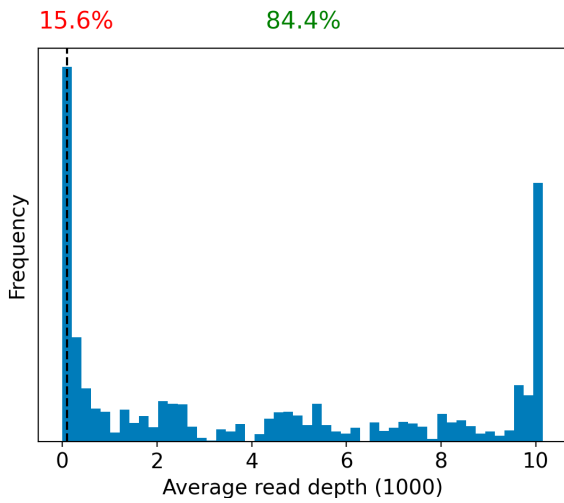
Sequencing chemistry:	Missing with Missing
Source site:	Missing (?.?)
Sampling date:	Missing
Collected by:	Missing
Sequenced by:	Missing
Total number of reads:	1115551
Reads aligned:	990608 (88%)
Average read quality:	19.8
Average read length:	696
Reads passing filter:	990602 (88%)
Average read quality passing filter:	19.8
Average read length passing filter:	696
Average coverage passing filter:	23056X

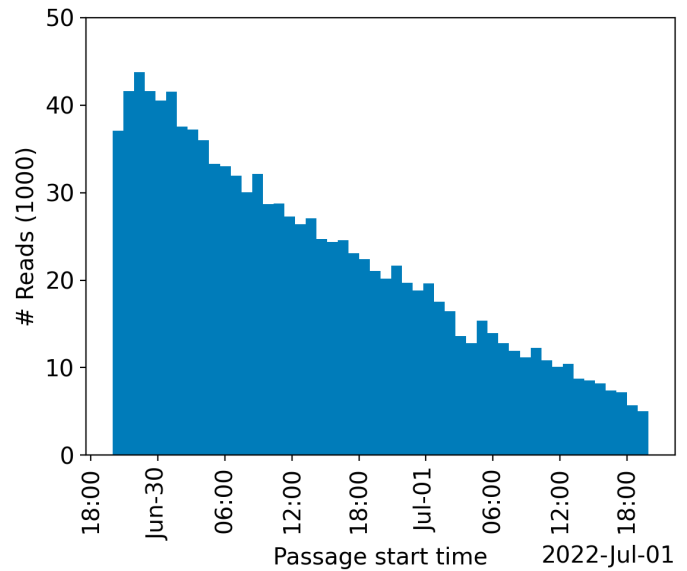
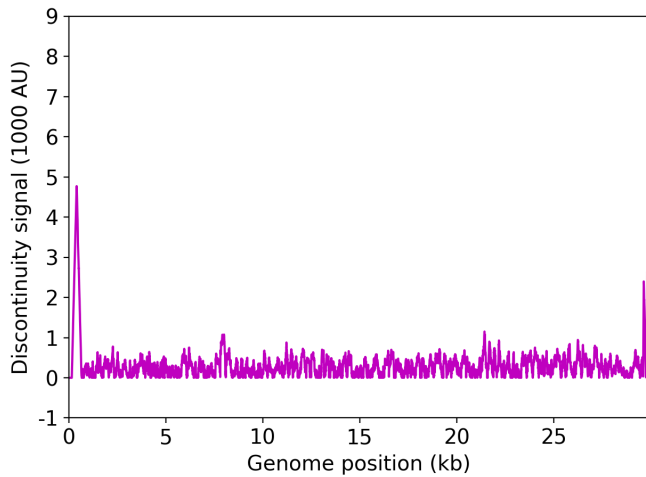
A read passes filter if the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4.

Overall sequence characteristics



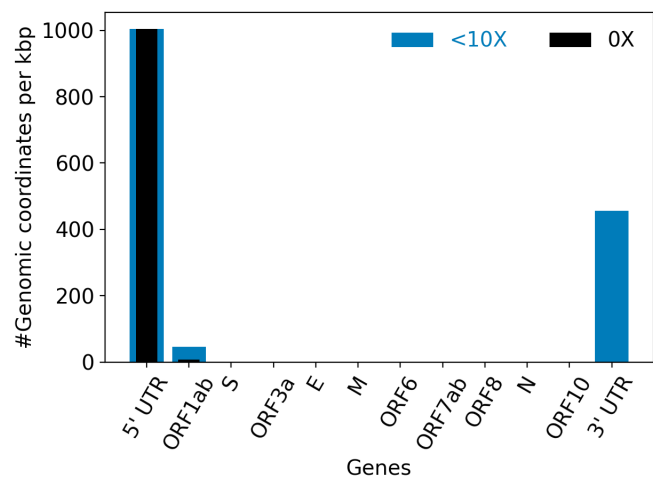
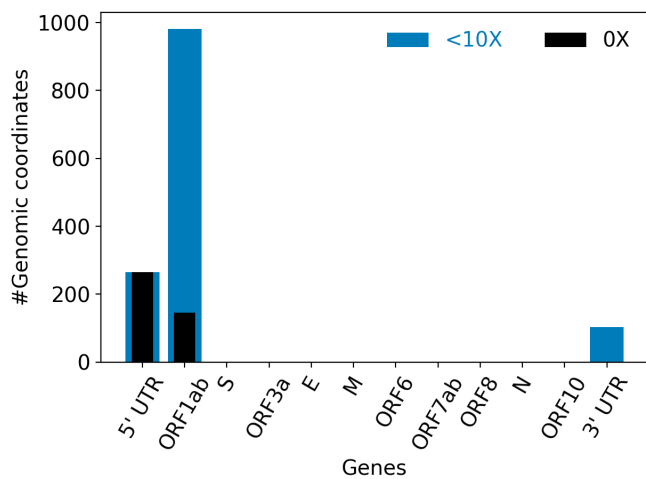
NOTE: The red shaded areas marked with a (*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.





	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	152nt (0%)	152nt (0%)
All genomic coordinates:	410nt (1%)	1350nt (4%)
Common SNPs:	1nt (2%)	1nt (2%)
Diverse SNPs:	2nt (0%)	30nt (13%)
Rare SNPs:	49nt (5%)	66nt (7%)

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.



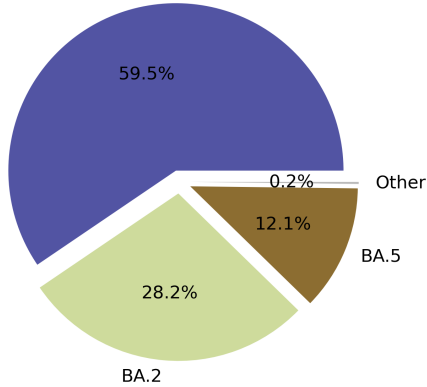
Hits to SARS-Cov2 genome (kraken2):	982438 reads (88.07%)
Hits to human genome (kraken2):	6838 reads (0.61%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
	Coronaviridae (88.07%)

Most abundant organisms (kraken2, family level):

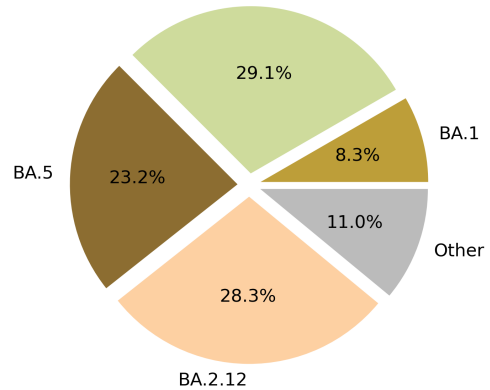
Arcobacteraceae (0.96%)
Hominidae (0.61%)
Enterobacteriaceae (0.16%)
Bacteroidaceae (0.12%)

Detected variants (Experimental)

Abundance of variants by linear regression
Omicron



Abundance of variants by kallisto
BA.2

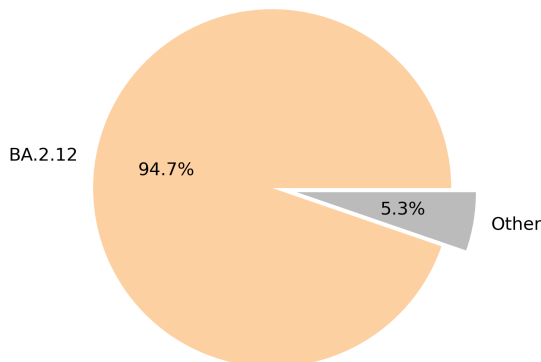


Based on deconvolution, [B.1.1.529](#) is estimated to constitute 59.47% of the viral particles and hence is the most abundant variant in the sample. The R^2 for the linear regression was 0.63. Variants that were detected less than 5% were grouped under "Other"

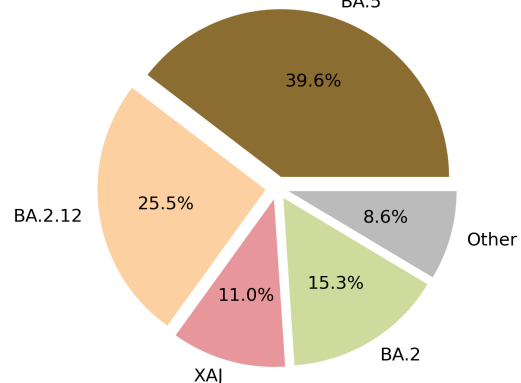
Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [BA.2](#) lineage. If this is a sample consisting of a single source of pathogens or an overwhelming majority of the different sources are infected with the same variant, the sample is dominated by this variant.

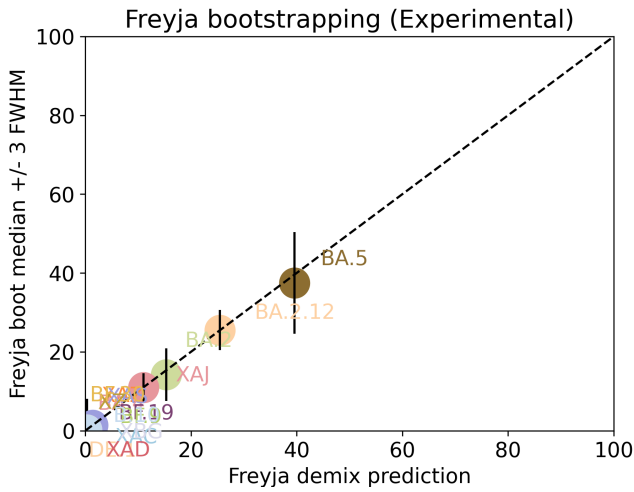
Based on mapping individual reads to the variant consensus sequences in the reference database, kallisto predicts that the sample is dominated by [BA.2](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.

Abundance of variants by kraken2+bracken, using majorCovid DB



Abundance of variants by Freyja
BA.5





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	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
Characteristic mutations detected	(3 of 13) S:G142D S:L452R S:T478K	(6 of 26) M:D3G NUC:C25000T NUC:C25584T ORF1AB:I3758V S:N856K S:Q493R	(23 of 31) N:S413R NUC:A20055G NUC:A9424G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:L3027F ORF1AB:L3201F ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:Q493R S:S371F S:R408S S:T19I S:T376A	(11 of 21) N:S413R NUC:C12880T NUC:C15714T NUC:C26858T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I S:D405N S:Q493R S:S371F	(22 of 31) N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A NUC:G27788T ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G	(20 of 28) N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G

[Jaccard Index](#) is a measure of similarity between two sets A and B, reaching the maximum value of 1 if A=B and minimum value of 0 if $A \cap B = \{\}$. In the c(d) representation below, c represents the Jaccard index of the set of mutations that were experimentally detected for this sample as listed above, whereas d refers to the ideal value of the Jaccard index expected from complete genome coverage without any sequencing errors.

	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
B.1.617.2	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.04 (0.02)	0.05 (0.03)
BA.1	0.00 (0.00)	1.00 (1.00)	0.12 (0.10)	0.06 (0.21)	0.08 (0.08)	0.08 (0.08)
BA.2	0.00 (0.00)	0.12 (0.10)	1.00 (1.00)	0.48 (0.33)	0.67 (0.63)	0.65 (0.59)
BA.3	0.00 (0.00)	0.06 (0.21)	0.48 (0.33)	1.00 (1.00)	0.43 (0.30)	0.41 (0.29)
BA.4	0.04 (0.02)	0.08 (0.08)	0.67 (0.63)	0.43 (0.30)	1.00 (1.00)	0.91 (0.84)

BA.5

0.05 ([0.03](#))

0.08 ([0.08](#))

0.65 ([0.59](#))

0.41 ([0.29](#))

0.91 ([0.84](#))

1.00 ([1.00](#))

Detected mutations

Excluded from this pdf version due to file size limitations.

CFSAN/OAO
BIostatistics AND BIOinformatics STAFF

WASTEWATER SARS-COV2 ANALYSIS REPORT

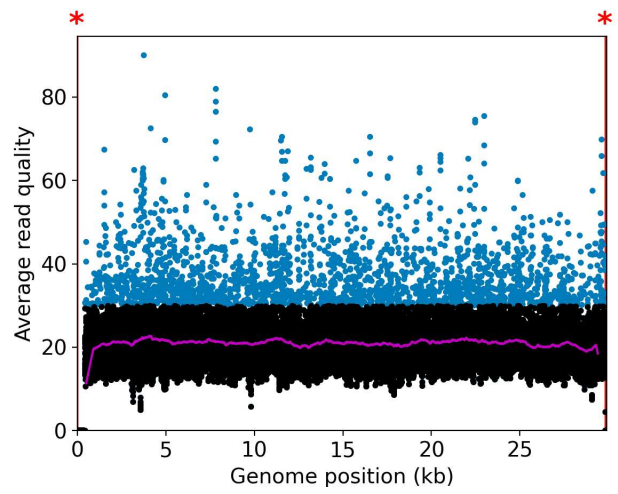
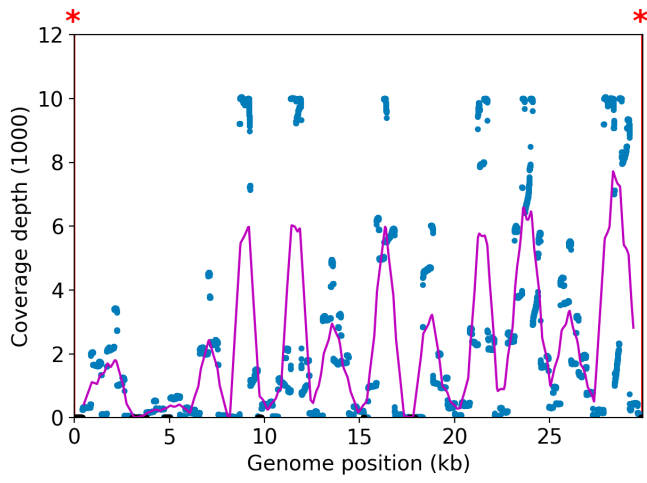
Sample name:	CFSANSMP000119225
Date generated:	2023-02-02, 09:10:20 EST
Timestamp of C-WAP version used:	Tue Jan 31 11:49:22 2023 -0500
Executed by:	Jasmine Amirzadegan (Jasmine.Amirzadegan@fda.hhs.gov)
Executed on:	172.20.44.125 (aka n125.raven.cfsan)

Sequencing summary

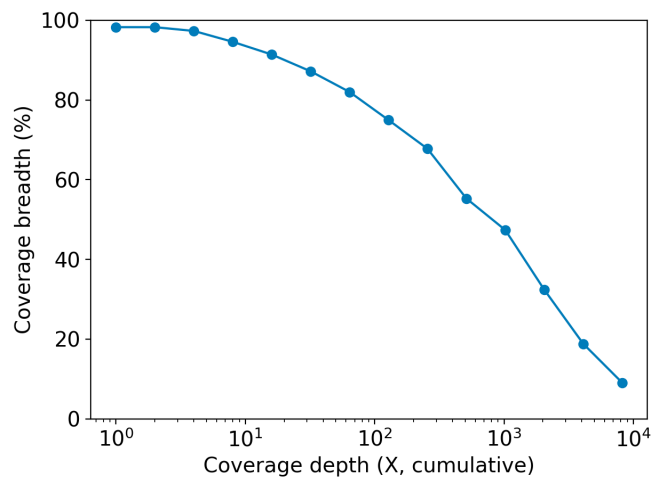
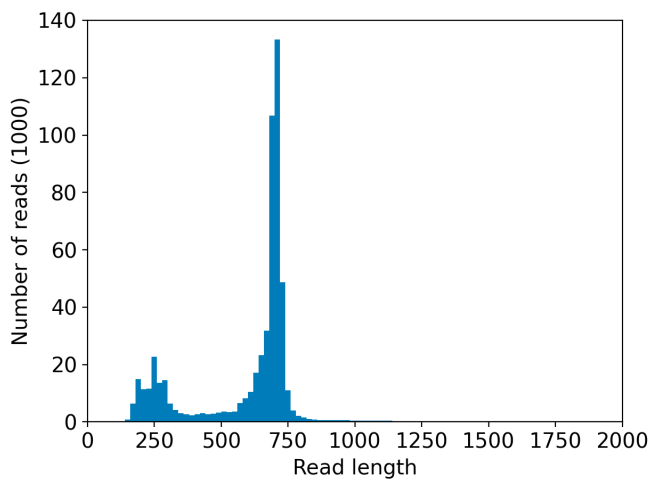
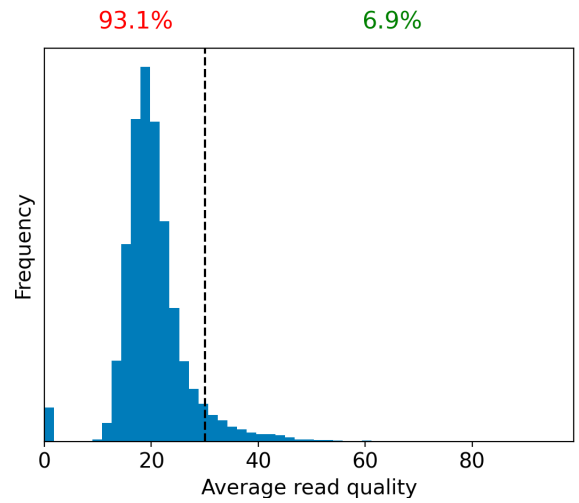
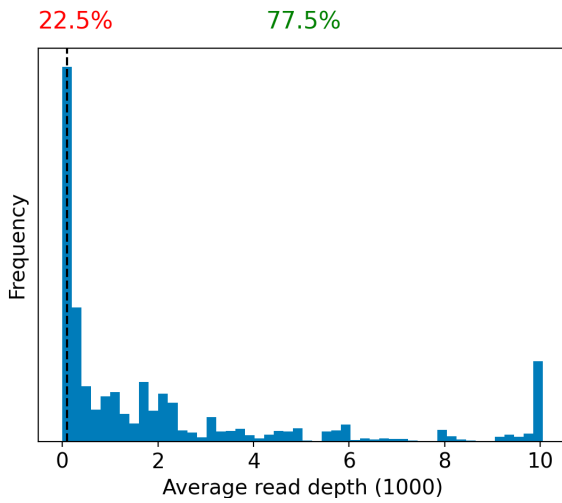
Sequencing chemistry:	Missing with Missing
Source site:	Missing (?.?)
Sampling date:	Missing
Collected by:	Missing
Sequenced by:	Missing
Total number of reads:	561738
Reads aligned:	448062 (79%)
Average read quality:	19.7
Average read length:	674
Reads passing filter:	448028 (79%)
Average read quality passing filter:	19.7
Average read length passing filter:	674
Average coverage passing filter:	10098X

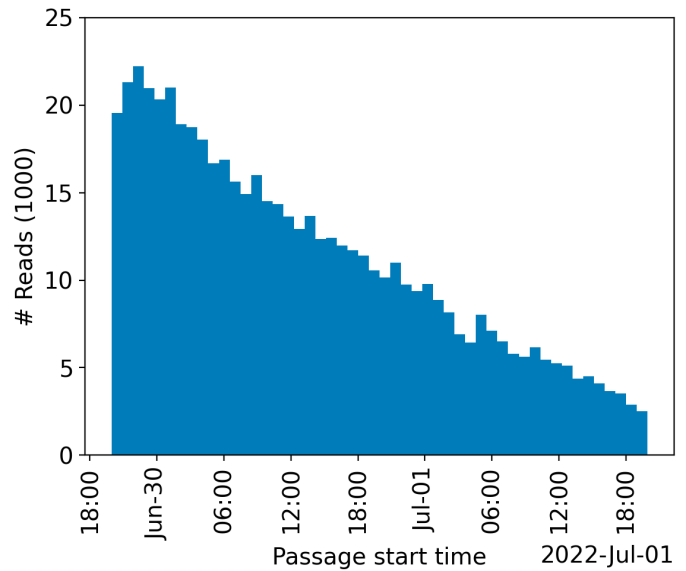
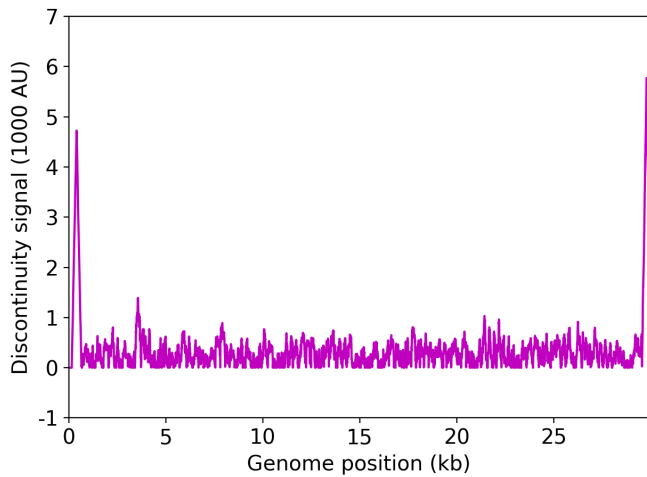
A read passes filter if the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4.

Overall sequence characteristics



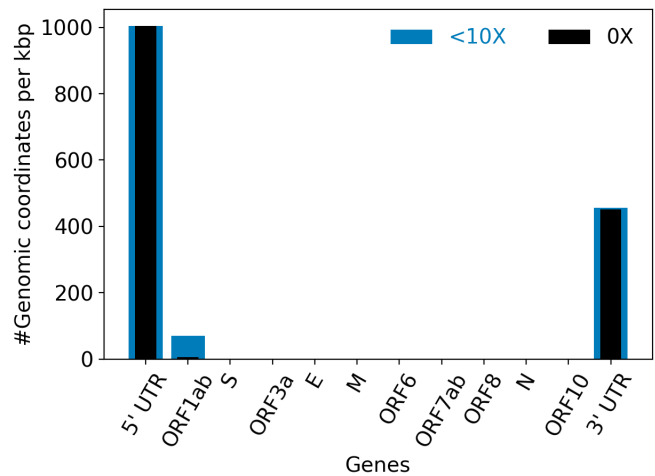
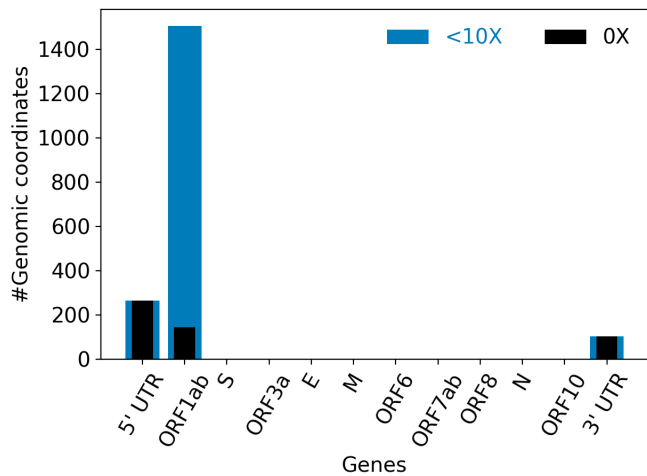
NOTE: The red shaded areas marked with a (*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.





	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	152nt (0%)	152nt (0%)
All genomic coordinates:	511nt (1%)	1874nt (6%)
Common SNPs:	1nt (2%)	1nt (2%)
Diverse SNPs:	29nt (12%)	29nt (12%)
Rare SNPs:	66nt (7%)	66nt (7%)

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.



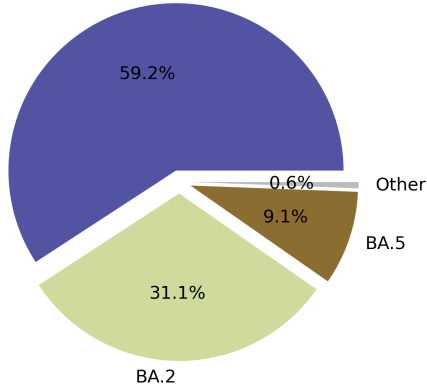
Hits to SARS-Cov2 genome (kraken2):	443611 reads (78.97%)
Hits to human genome (kraken2):	5188 reads (0.92%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
	Coronaviridae (78.97%)

Most abundant organisms (kraken2, family level):

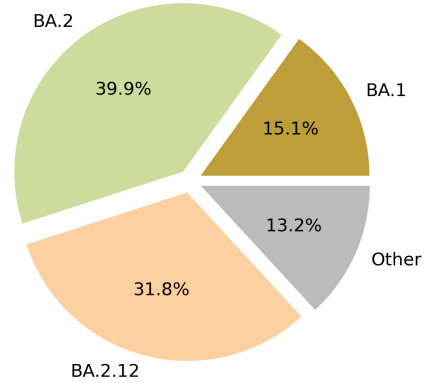
Arcobacteraceae (1.92%)
Hominidae (0.92%)
Enterobacteriaceae (0.32%)
Bacteroidaceae (0.27%)

Detected variants (Experimental)

Abundance of variants by linear regression
Omicron



Abundance of variants by kallisto

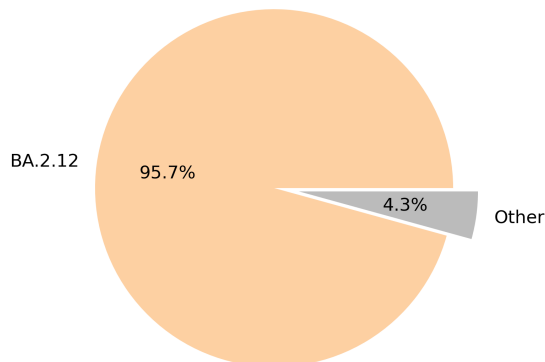


Based on deconvolution, [B.1.1.529](#) is estimated to constitute 59.23% of the viral particles and hence is the most abundant variant in the sample. The R^2 for the linear regression was 0.63. Variants that were detected less than 5% were grouped under "Other"

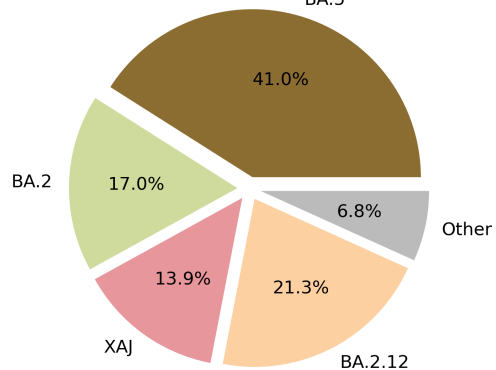
Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [BA.2](#) lineage. If this is a sample consisting of a single source of pathogens or an overwhelming majority of the different sources are infected with the same variant, the sample is dominated by this variant.

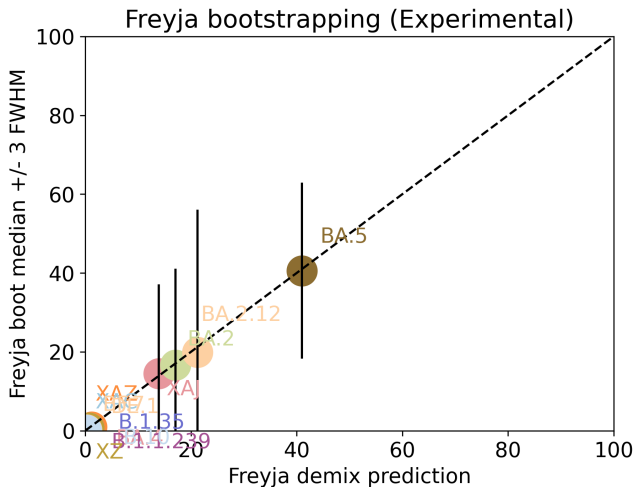
Based on mapping individual reads to the variant consensus sequences in the reference database, kallisto predicts that the sample is dominated by [BA.2](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.

Abundance of variants by kraken2+bracken, using majorCovid DB



Abundance of variants by Freyja
BA.5





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	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
Characteristic mutations detected	(3 of 13) S:G142D S:L452R S:T478K	(5 of 26) M:D3G NUC:C25000T ORF1AB:I3758V S:Q493R	(23 of 31) N:S413R NUC:A20055G NUC:A9424G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:L3027F ORF1AB:L3201F ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:Q493R S:S371F S:R408S S:T19I S:T376A	(11 of 21) N:S413R NUC:C12880T NUC:C15714T NUC:C26858T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I S:D405N S:Q493R S:S371F	(22 of 31) N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A NUC:G27788T ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G	(21 of 28) M:D3N N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G

[Jaccard Index](#) is a measure of similarity between two sets A and B, reaching the maximum value of 1 if $A=B$ and minimum value of 0 if $A \cap B = \{\}$. In the c(d) representation below, c represents the Jaccard index of the set of mutations that were experimentally detected for this sample as listed above, whereas d refers to the ideal value of the Jaccard index expected from complete genome coverage without any sequencing errors.

	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
B.1.617.2	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.04 (0.02)	0.04 (0.03)
BA.1	0.00 (0.00)	1.00 (1.00)	0.12 (0.10)	0.07 (0.21)	0.08 (0.08)	0.08 (0.08)
BA.2	0.00 (0.00)	0.12 (0.10)	1.00 (1.00)	0.48 (0.33)	0.67 (0.63)	0.63 (0.59)
BA.3	0.00 (0.00)	0.07 (0.21)	0.48 (0.33)	1.00 (1.00)	0.43 (0.30)	0.39 (0.29)
BA.4	0.04 (0.02)	0.08 (0.08)	0.67 (0.63)	0.43 (0.30)	1.00 (1.00)	0.87 (0.84)

BA.5

0.04 ([0.03](#))

0.08 ([0.08](#))

0.63 ([0.59](#))

0.39 ([0.29](#))

0.87 ([0.84](#))

1.00 ([1.00](#))

Detected mutations

Excluded from this pdf version due to file size limitations.

CFSAN/OAO
BIostatistics AND Bioinformatics Staff

WASTEWATER SARS-COV2 ANALYSIS REPORT

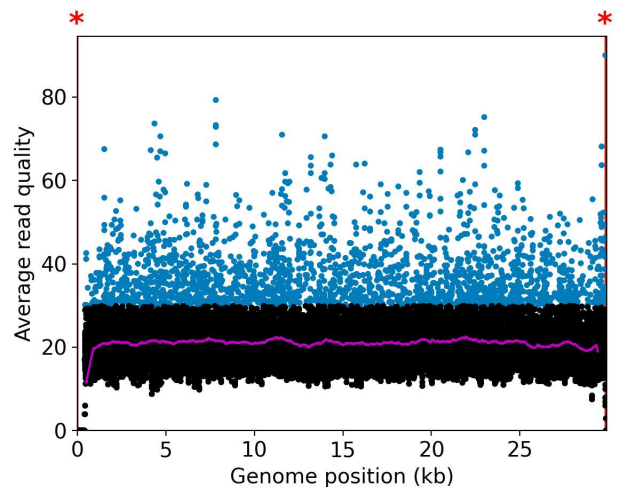
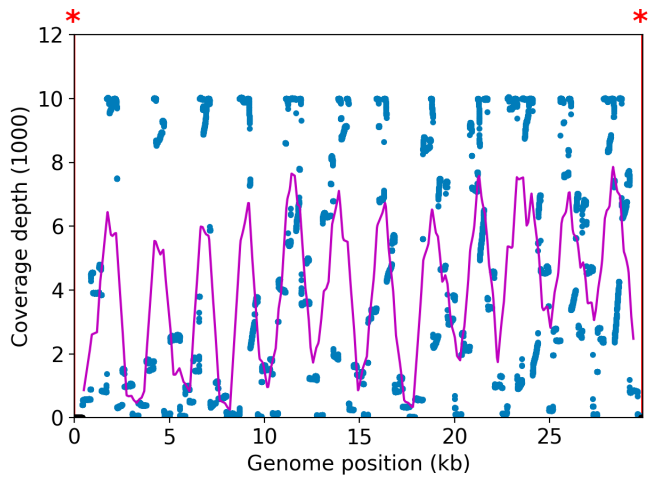
Sample name:	CFSANSMP000119226
Date generated:	2023-02-02, 09:10:42 EST
Timestamp of C-WAP version used:	Tue Jan 31 11:49:22 2023 -0500
Executed by:	Jasmine Amirzadegan (Jasmine.Amirzadegan@fda.hhs.gov)
Executed on:	172.20.44.140 (aka n140.raven.cfsan)

Sequencing summary

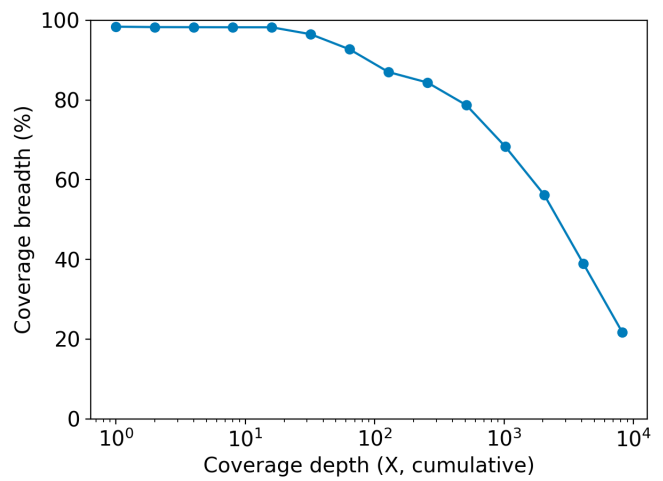
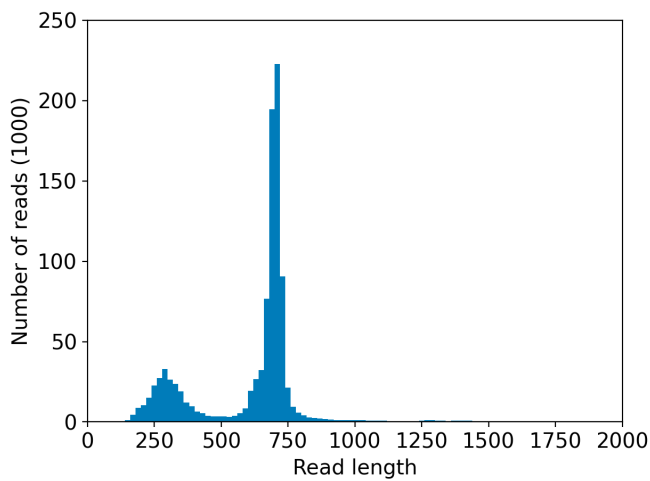
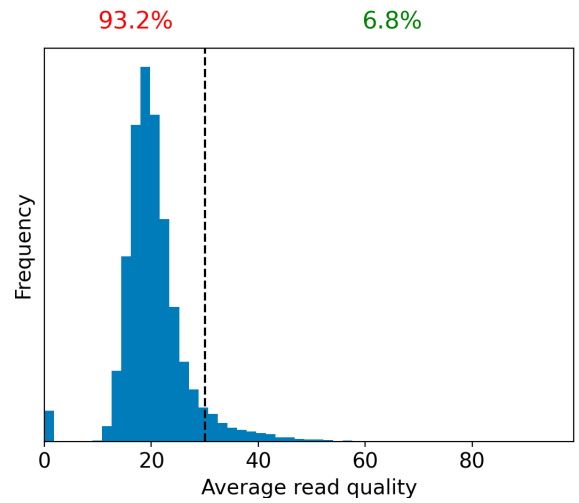
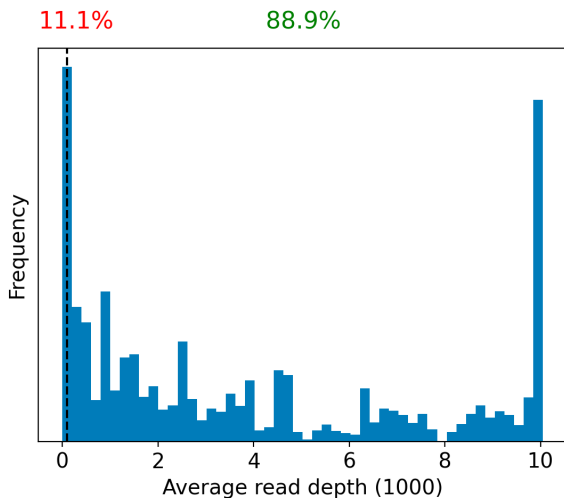
Sequencing chemistry:	Missing with Missing
Source site:	Missing (?.?)
Sampling date:	Missing
Collected by:	Missing
Sequenced by:	Missing
Total number of reads:	1638438
Reads aligned:	1286637 (78%)
Average read quality:	19.8
Average read length:	692
Reads passing filter:	1286629 (78%)
Average read quality passing filter:	19.8
Average read length passing filter:	692
Average coverage passing filter:	29774X

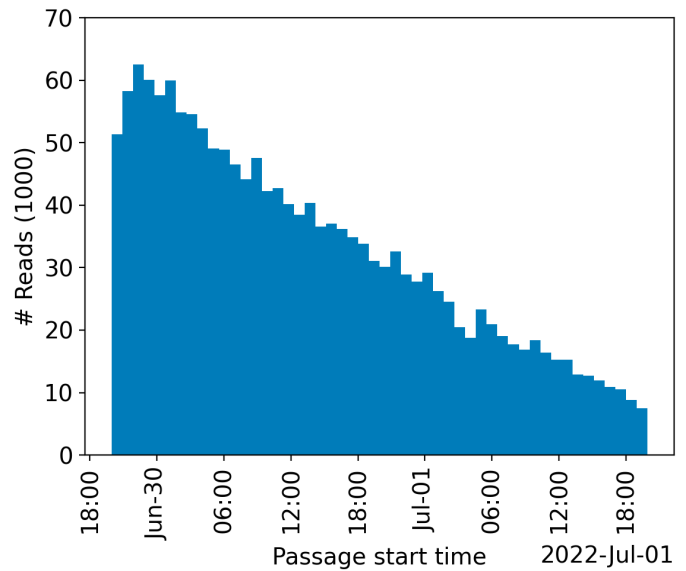
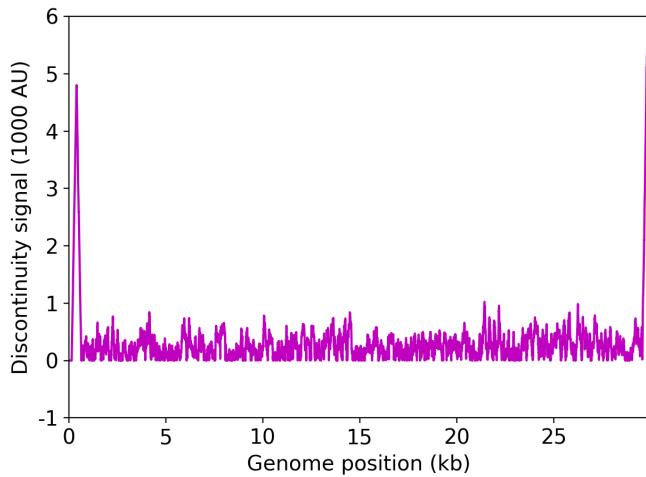
A read passes filter if the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4.

Overall sequence characteristics



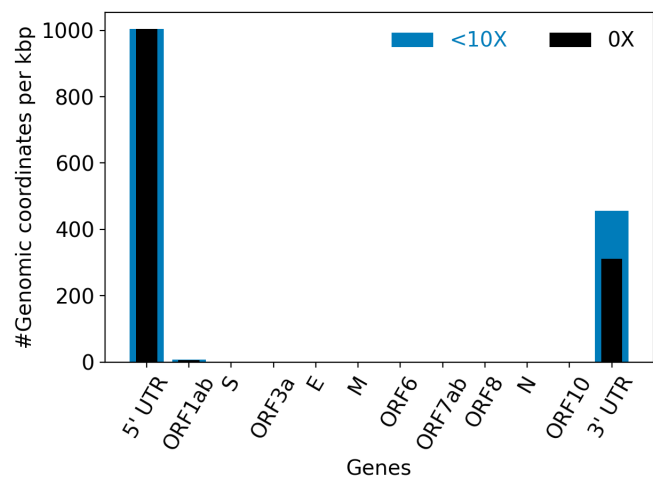
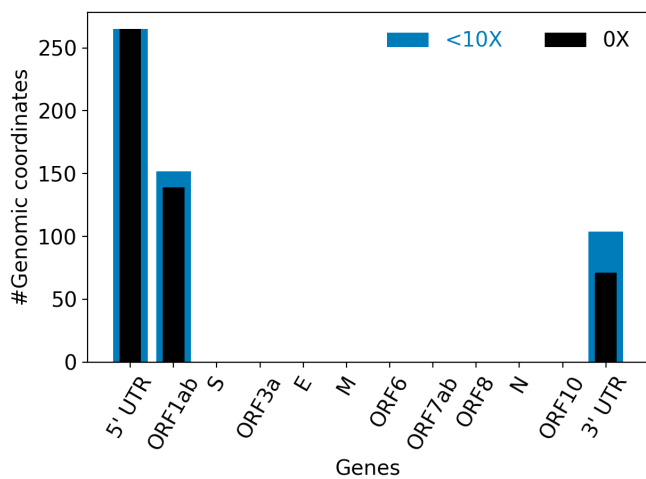
NOTE: The red shaded areas marked with a (*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.





	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	152nt (0%)	152nt (0%)
All genomic coordinates:	475nt (1%)	521nt (1%)
Common SNPs:	1nt (2%)	1nt (2%)
Diverse SNPs:	29nt (12%)	29nt (12%)
Rare SNPs:	61nt (6%)	66nt (7%)

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.



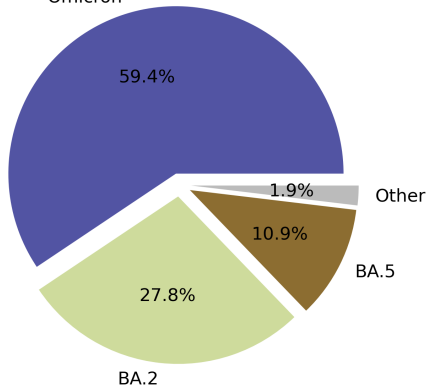
Hits to SARS-Cov2 genome (kraken2):	1276934 reads (77.94%)
Hits to human genome (kraken2):	13346 reads (0.81%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
	Coronaviridae (77.94%)

Most abundant organisms (kraken2, family level):

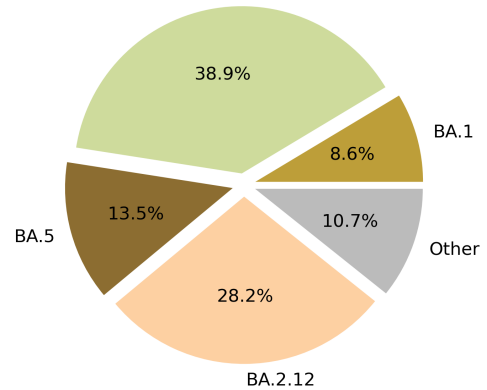
Arcobacteraceae (1.39%)
Hominidae (0.81%)
Enterobacteriaceae (0.21%)
Bacteroidaceae (0.09%)

Detected variants (Experimental)

Abundance of variants by linear regression
Omicron



Abundance of variants by kallisto
BA.2

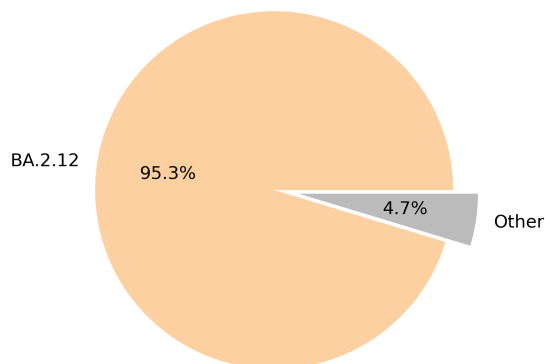


Based on deconvolution, [B.1.1.529](#) is estimated to constitute 59.36% of the viral particles and hence is the most abundant variant in the sample. The R^2 for the linear regression was 0.63. Variants that were detected less than 5% were grouped under "Other"

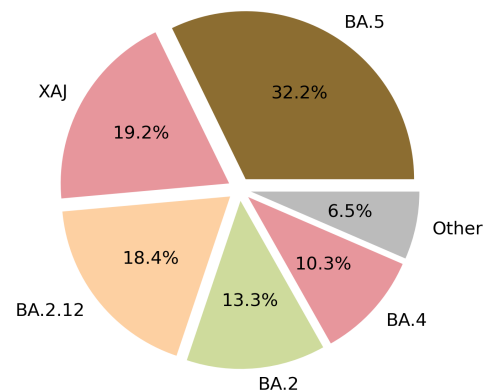
Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [BA.2](#) lineage. If this is a sample consisting of a single source of pathogens or an overwhelming majority of the different sources are infected with the same variant, the sample is dominated by this variant.

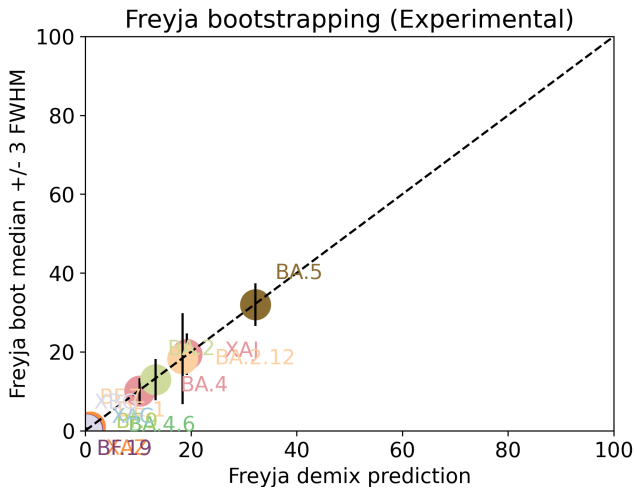
Based on mapping individual reads to the variant consensus sequences in the reference database, kallisto predicts that the sample is dominated by [BA.2](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.

Abundance of variants by kraken2+bracken, using majorCovid DB



Abundance of variants by Freyja





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	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
Characteristic mutations detected	(3 of 13) S:G142D S:L452R S:T478K	(5 of 26) M:D3G NUC:C25000T NUC:C25584T ORF1AB:I3758V S:Q493R	(23 of 31) N:S413R NUC:A20055G NUC:A9424G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:L3027F ORF1AB:L3201F ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:Q493R S:S371F S:R408S S:T19I S:T376A	(11 of 21) N:S413R NUC:C12880T NUC:C15714T NUC:C26858T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I S:D405N S:Q493R S:S371F	(22 of 31) N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A NUC:G27788T ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G	(21 of 28) M:D3N N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G

[Jaccard Index](#) is a measure of similarity between two sets A and B, reaching the maximum value of 1 if A=B and minimum value of 0 if $A \cap B = \{\}$. In the c(d) representation below, c represents the Jaccard index of the set of mutations that were experimentally detected for this sample as listed above, whereas d refers to the ideal value of the Jaccard index expected from complete genome coverage without any sequencing errors.

	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
B.1.617.2	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.04 (0.02)	0.04 (0.03)
BA.1	0.00 (0.00)	1.00 (1.00)	0.12 (0.10)	0.07 (0.21)	0.08 (0.08)	0.08 (0.08)
BA.2	0.00 (0.00)	0.12 (0.10)	1.00 (1.00)	0.48 (0.33)	0.67 (0.63)	0.63 (0.59)
BA.3	0.00 (0.00)	0.07 (0.21)	0.48 (0.33)	1.00 (1.00)	0.43 (0.30)	0.39 (0.29)
BA.4	0.04 (0.02)	0.08 (0.08)	0.67 (0.63)	0.43 (0.30)	1.00 (1.00)	0.87 (0.84)

BA.5

0.04 ([0.03](#))

0.08 ([0.08](#))

0.63 ([0.59](#))

0.39 ([0.29](#))

0.87 ([0.84](#))

1.00 ([1.00](#))

Detected mutations

Excluded from this pdf version due to file size limitations.

CFSAN/OAO
BIostatistics AND BIOinformatics STAFF

WASTEWATER SARS-COV2 ANALYSIS REPORT

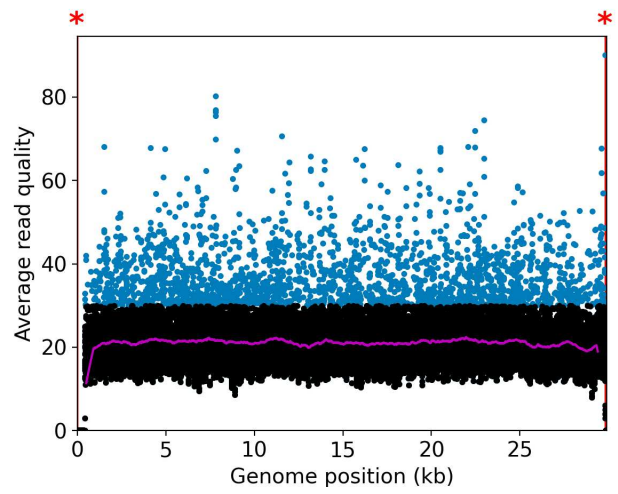
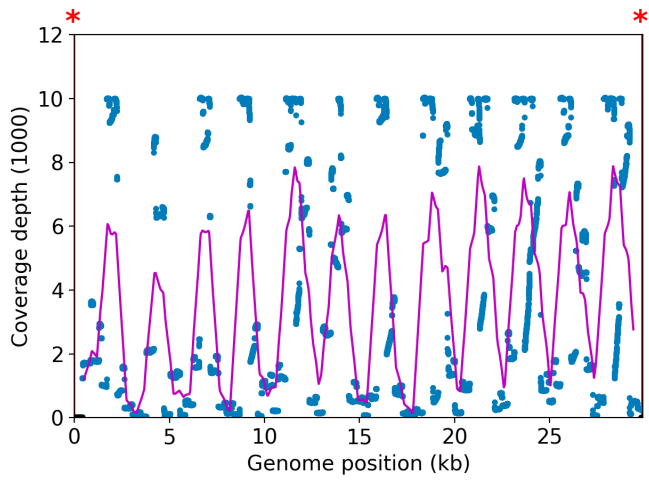
Sample name:	CFSANSMP000119227
Date generated:	2023-02-02, 09:08:40 EST
Timestamp of C-WAP version used:	Tue Jan 31 11:49:22 2023 -0500
Executed by:	Jasmine Amirzadegan (Jasmine.Amirzadegan@fda.hhs.gov)
Executed on:	172.20.44.123 (aka n123.raven.cfsan)

Sequencing summary

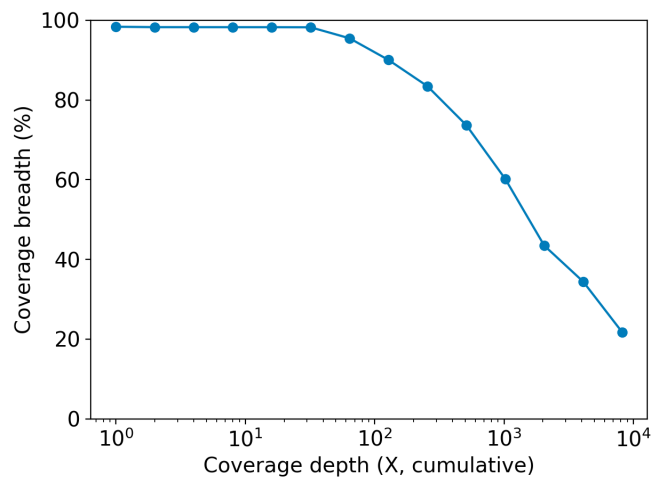
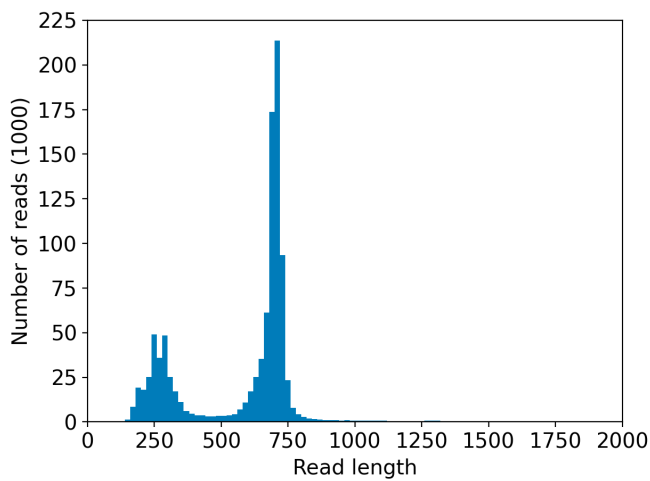
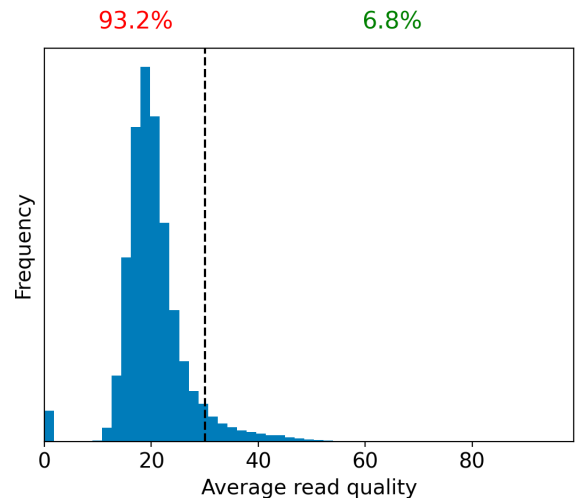
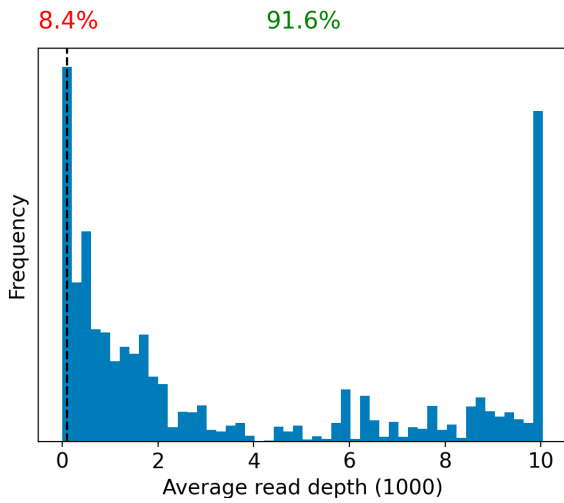
Sequencing chemistry:	Missing with Missing
Source site:	Missing (?.?)
Sampling date:	Missing
Collected by:	Missing
Sequenced by:	Missing
Total number of reads:	1336574
Reads aligned:	1005999 (75%)
Average read quality:	19.7
Average read length:	679
Reads passing filter:	1005993 (75%)
Average read quality passing filter:	19.7
Average read length passing filter:	679
Average coverage passing filter:	22842X

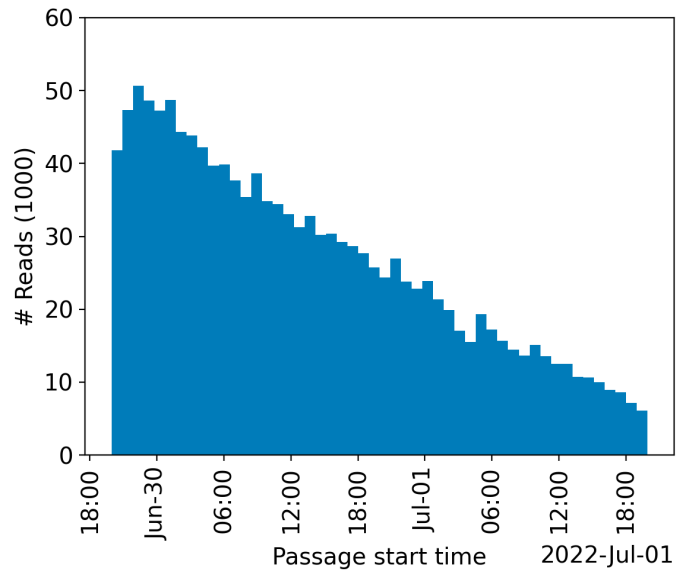
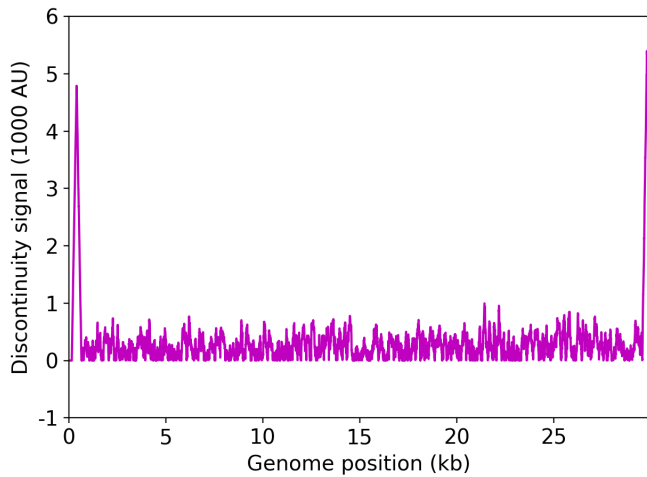
A read passes filter if the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4.

Overall sequence characteristics



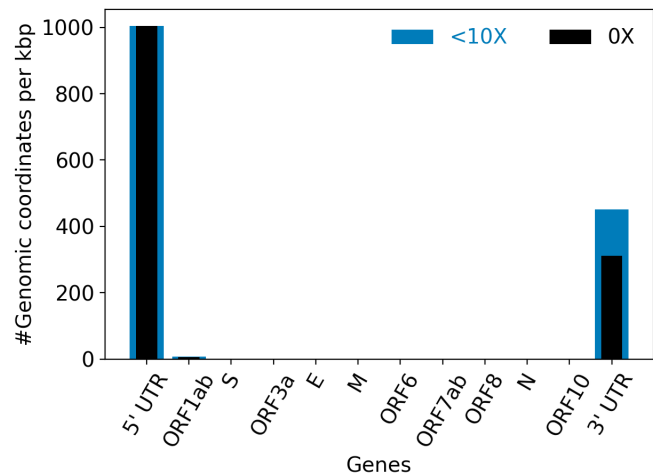
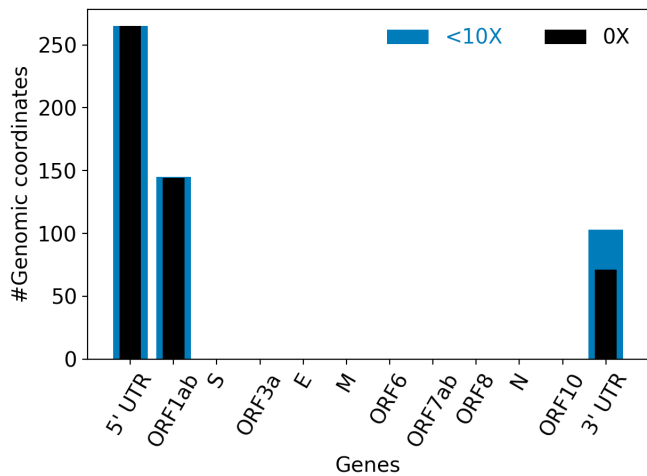
NOTE: The red shaded areas marked with a (*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.





	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	152nt (0%)	152nt (0%)
All genomic coordinates:	480nt (1%)	513nt (1%)
Common SNPs:	1nt (2%)	1nt (2%)
Diverse SNPs:	29nt (12%)	29nt (12%)
Rare SNPs:	61nt (6%)	66nt (7%)

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.



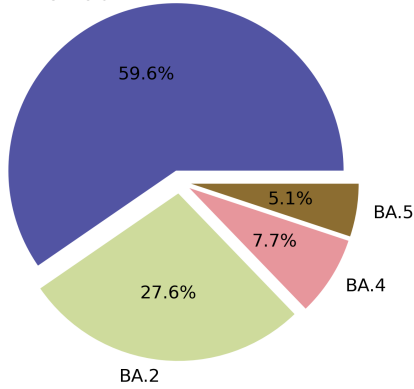
Hits to SARS-Cov2 genome (kraken2):	996786 reads (74.58%)
Hits to human genome (kraken2):	12251 reads (0.92%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
	Coronaviridae (74.58%)

Most abundant organisms (kraken2, family level):

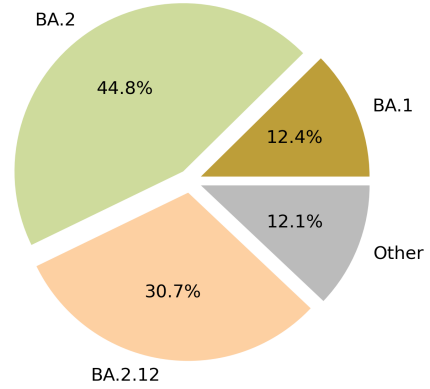
Arcobacteraceae (1.18%)
Hominidae (0.92%)
Enterobacteriaceae (0.25%)
Mycobacteriaceae (0.15%)

Detected variants (Experimental)

Abundance of variants by linear regression
Omicron



Abundance of variants by kallisto

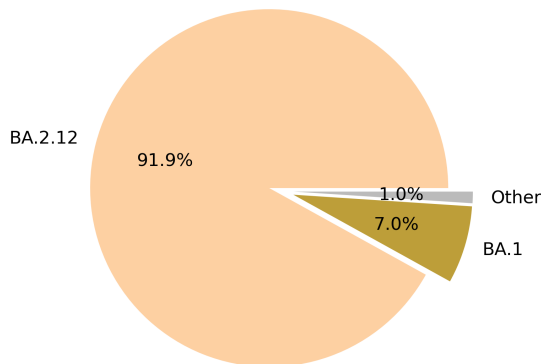


Based on deconvolution, [B.1.1.529](#) is estimated to constitute 59.58% of the viral particles and hence is the most abundant variant in the sample. The R^2 for the linear regression was 0.63. Variants that were detected less than 5% were grouped under "Other"

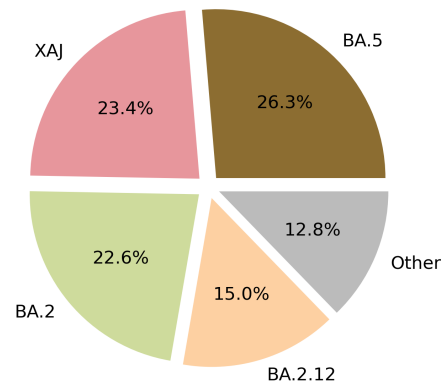
Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [BA.2](#) lineage. If this is a sample consisting of a single source of pathogens or an overwhelming majority of the different sources are infected with the same variant, the sample is dominated by this variant.

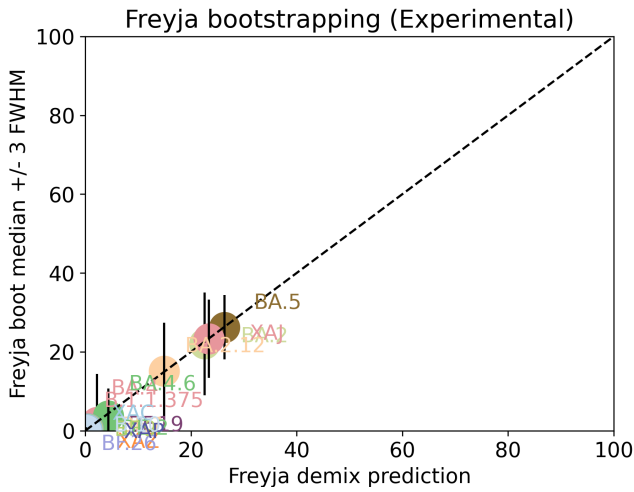
Based on mapping individual reads to the variant consensus sequences in the reference database, kallisto predicts that the sample is dominated by [BA.2](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.

Abundance of variants by kraken2+bracken, using majorCovid DB



Abundance of variants by Freyja





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	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
Characteristic mutations detected	(3 of 13) S:G142D S:L452R S:T478K	(6 of 26) M:D3G NUC:C25000T NUC:C25584T NUC:T13195C ORF1AB:I3758V S:Q493R	(23 of 31) N:S413R NUC:A20055G NUC:A9424G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:L3027F ORF1AB:L3201F ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:Q493R S:S371F S:R408S S:T19I S:T376A	(11 of 21) N:S413R NUC:C12880T NUC:C15714T NUC:C26858T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I S:D405N S:Q493R S:S371F	(22 of 31) N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A NUC:G27788T ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G	(20 of 28) N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G

[Jaccard Index](#) is a measure of similarity between two sets A and B, reaching the maximum value of 1 if A=B and minimum value of 0 if $A \cap B = \{\}$. In the c(d) representation below, c represents the Jaccard index of the set of mutations that were experimentally detected for this sample as listed above, whereas d refers to the ideal value of the Jaccard index expected from complete genome coverage without any sequencing errors.

	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
B.1.617.2	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.04 (0.02)	0.05 (0.03)
BA.1	0.00 (0.00)	1.00 (1.00)	0.12 (0.10)	0.06 (0.21)	0.08 (0.08)	0.08 (0.08)
BA.2	0.00 (0.00)	0.12 (0.10)	1.00 (1.00)	0.48 (0.33)	0.67 (0.63)	0.65 (0.59)
BA.3	0.00 (0.00)	0.06 (0.21)	0.48 (0.33)	1.00 (1.00)	0.43 (0.30)	0.41 (0.29)
BA.4	0.04 (0.02)	0.08 (0.08)	0.67 (0.63)	0.43 (0.30)	1.00 (1.00)	0.91 (0.84)

BA.5

0.05 ([0.03](#))

0.08 ([0.08](#))

0.65 ([0.59](#))

0.41 ([0.29](#))

0.91 ([0.84](#))

1.00 ([1.00](#))

Detected mutations

Excluded from this pdf version due to file size limitations.

**CFSAN/OAO
BIostatistics AND BIOinformatics STAFF**

WASTEWATER SARS-COV2 ANALYSIS REPORT

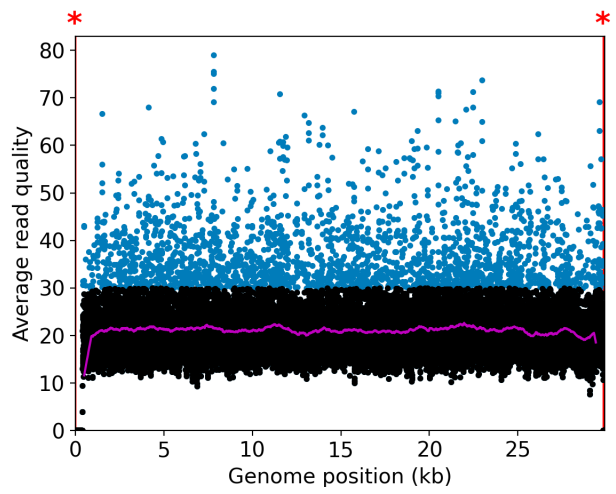
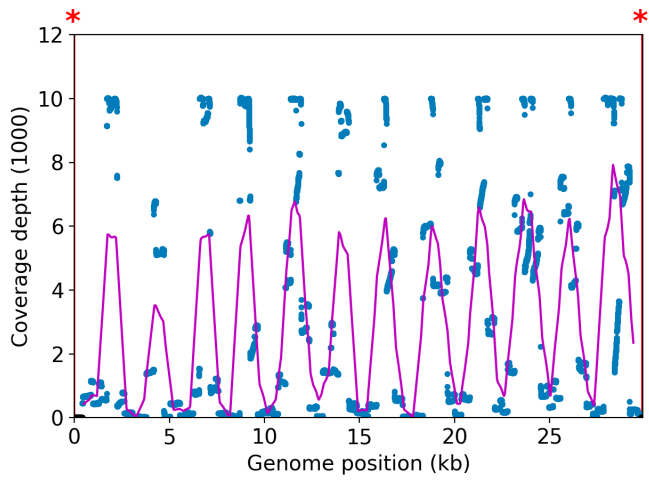
Sample name:	CFSANSMP000119228
Date generated:	2023-02-02, 09:08:23 EST
Timestamp of C-WAP version used:	Tue Jan 31 11:49:22 2023 -0500
Executed by:	Jasmine Amirzadegan (jasmine.Amirzadegan@fda.hhs.gov)
Executed on:	172.20.44.123 (aka n123.raven.cfsan)

Sequencing summary

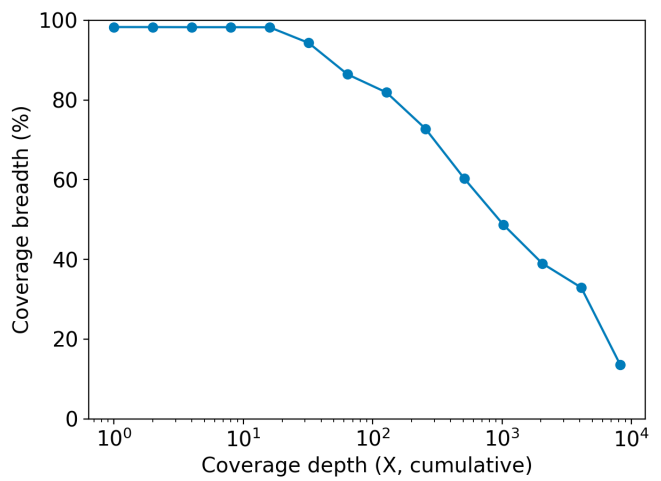
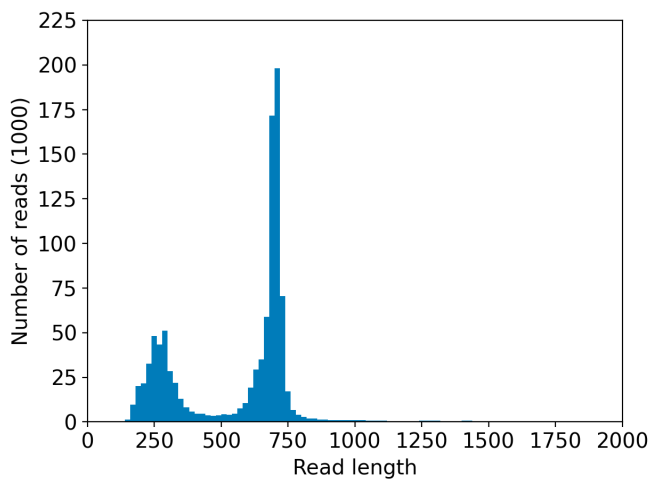
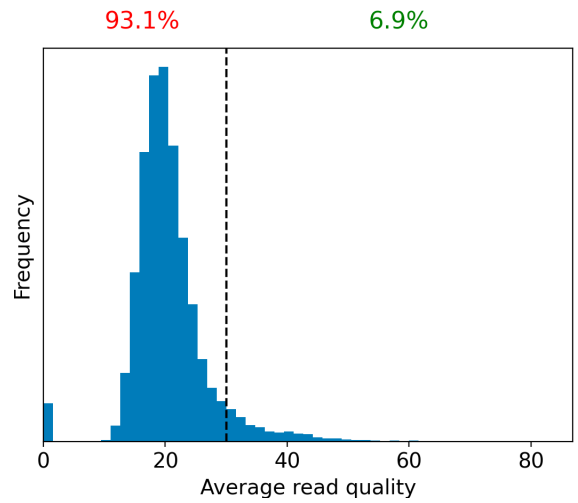
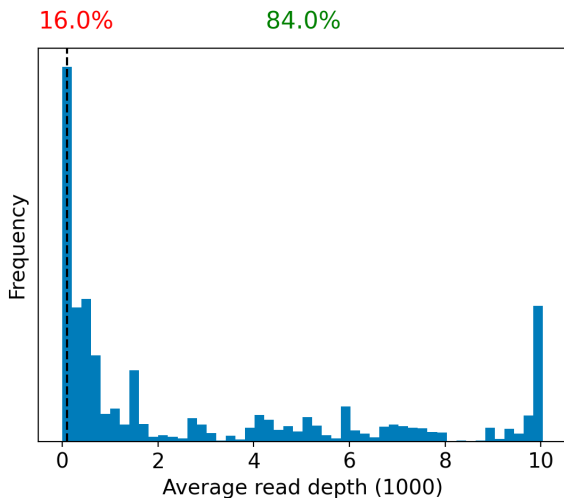
Sequencing chemistry:	Missing with Missing
Source site:	Missing (?.?)
Sampling date:	Missing
Collected by:	Missing
Sequenced by:	Missing
Total number of reads:	1037901
Reads aligned:	725757 (69%)
Average read quality:	19.7
Average read length:	681
Reads passing filter:	725752 (69%)
Average read quality passing filter:	19.7
Average read length passing filter:	681
Average coverage passing filter:	16528X

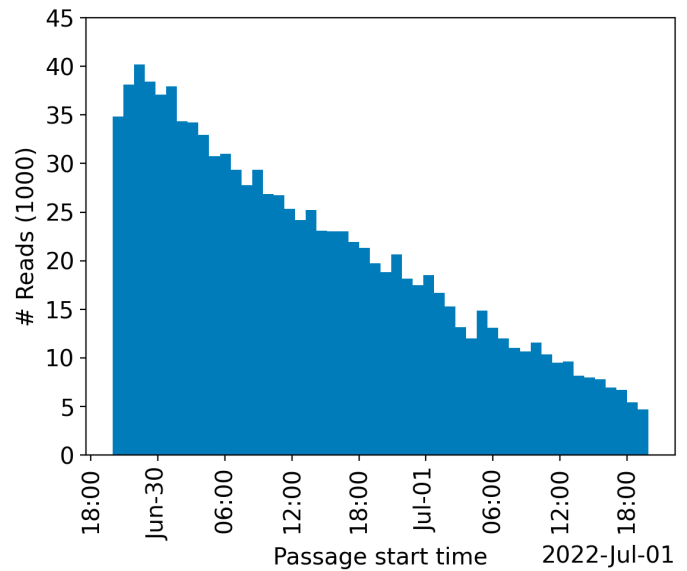
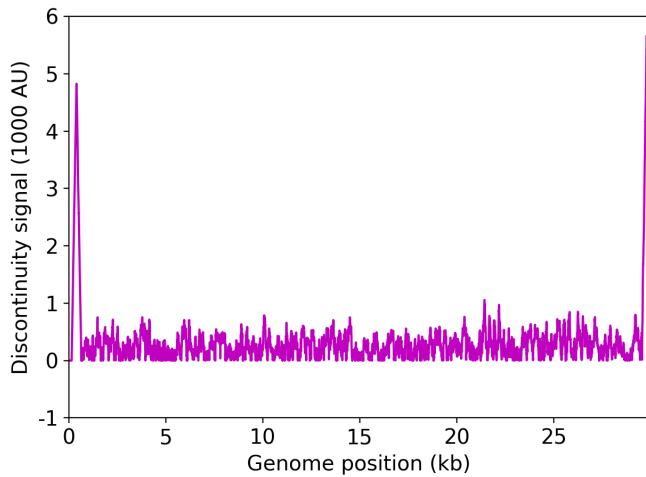
A read passes filter if the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4.

Overall sequence characteristics



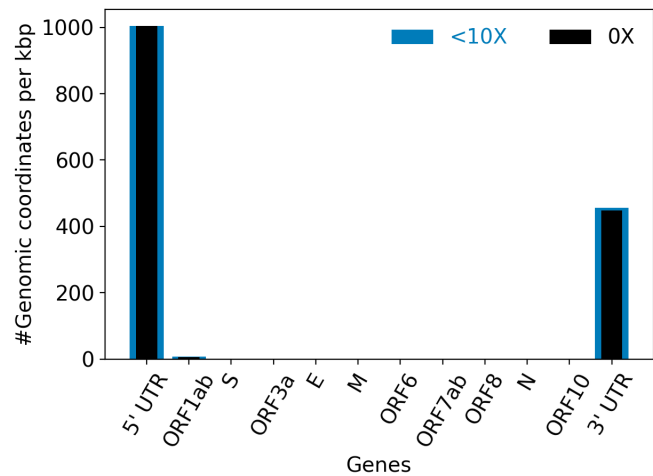
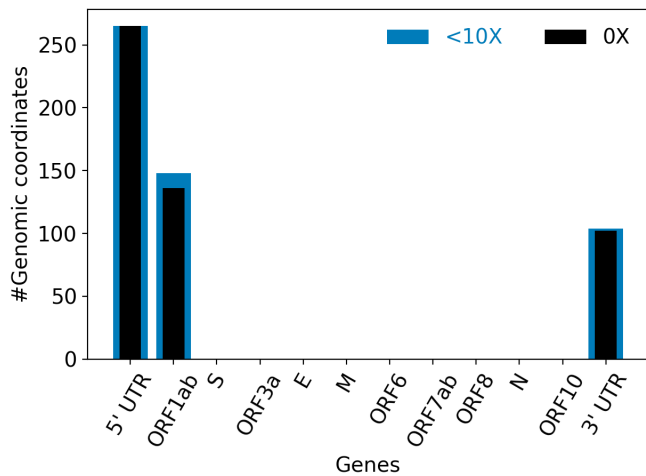
NOTE: The red shaded areas marked with a (*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.





	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	152nt (0%)	152nt (0%)
All genomic coordinates:	503nt (1%)	517nt (1%)
Common SNPs:	1nt (2%)	1nt (2%)
Diverse SNPs:	29nt (12%)	29nt (12%)
Rare SNPs:	66nt (7%)	66nt (7%)

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.



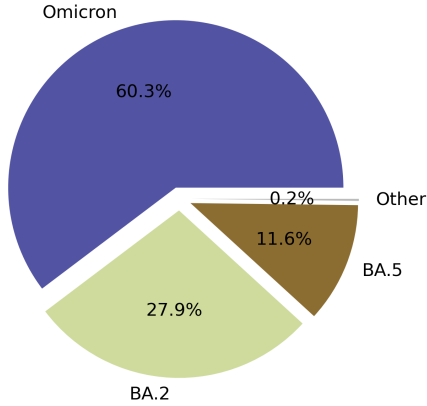
Hits to SARS-Cov2 genome (kraken2):	718697 reads (69.25%)
Hits to human genome (kraken2):	10856 reads (1.05%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
	Coronaviridae (69.25%)

Most abundant organisms (kraken2, family level):

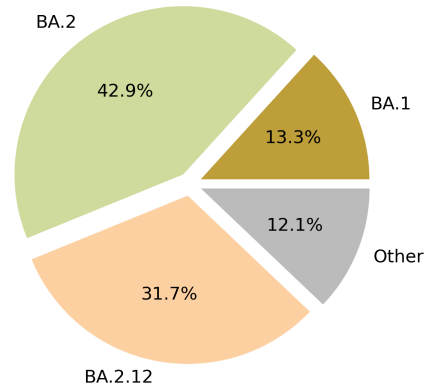
Enterobacteriaceae (7.07%)
Arcobacteraceae (1.61%)
Hominidae (1.05%)
Bacteroidaceae (0.14%)

Detected variants (Experimental)

Abundance of variants by linear regression



Abundance of variants by kallisto

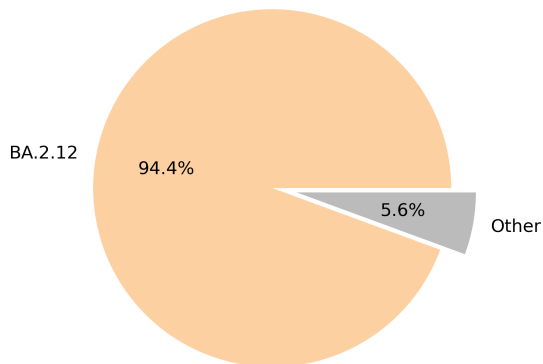


Based on deconvolution, [B.1.1.529](#) is estimated to constitute 60.35% of the viral particles and hence is the most abundant variant in the sample. The R^2 for the linear regression was 0.62. Variants that were detected less than 5% were grouped under "Other"

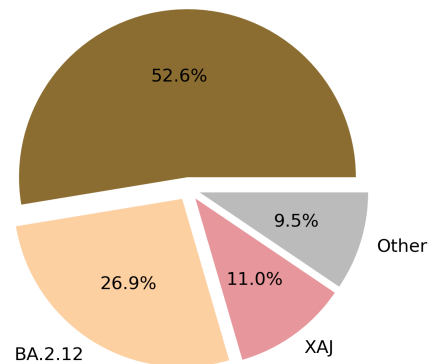
Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [BA.2](#) lineage. If this is a sample consisting of a single source of pathogens or an overwhelming majority of the different sources are infected with the same variant, the sample is dominated by this variant.

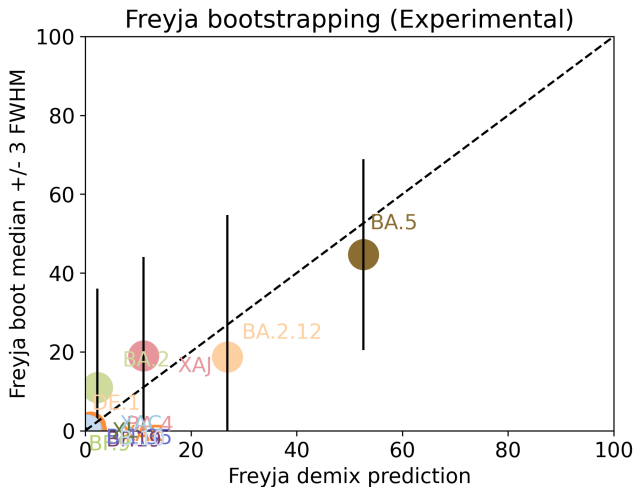
Based on mapping individual reads to the variant consensus sequences in the reference database, kallisto predicts that the sample is dominated by [BA.2](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.

Abundance of variants by kraken2+bracken, using majorCovid DB



Abundance of variants by Freyja





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	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
Characteristic mutations detected	(3 of 13) S:G142D S:L452R S:T478K	(5 of 26) M:D3G NUC:C25000T NUC:C25584T ORF1AB:I3758V S:Q493R	(22 of 31) N:S413R NUC:A9424G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:L3027F ORF1AB:L3201F ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:Q493R S:R408S S:S371F S:T19I S:T376A	(11 of 21) N:S413R NUC:C12880T NUC:C15714T NUC:C26858T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I S:D405N S:Q493R S:S371F	(20 of 31) N:S413R NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G	(20 of 28) M:D3N N:S413R NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G

Jaccard Index is a measure of similarity between two sets A and B, reaching the maximum value of 1 if A=B and minimum value of 0 if $A \cap B = \{\}$. In the c(d) representation below, c represents the Jaccard index of the set of mutations that were experimentally detected for this sample as listed above, whereas d refers to the ideal value of the Jaccard index expected from complete genome coverage without any sequencing errors.

	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
B.1.617.2	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.05 (0.02)	0.05 (0.03)
BA.1	0.00 (0.00)	1.00 (1.00)	0.12 (0.10)	0.07 (0.21)	0.09 (0.08)	0.09 (0.08)
BA.2	0.00 (0.00)	0.12 (0.10)	1.00 (1.00)	0.50 (0.33)	0.68 (0.63)	0.62 (0.59)
BA.3	0.00 (0.00)	0.07 (0.21)	0.50 (0.33)	1.00 (1.00)	0.48 (0.30)	0.41 (0.29)
BA.4	0.05 (0.02)	0.09 (0.08)	0.68 (0.63)	0.48 (0.30)	1.00 (1.00)	0.90 (0.84)

BA.5

0.05 ([0.03](#))

0.09 ([0.08](#))

0.62 ([0.59](#))

0.41 ([0.29](#))

0.90 ([0.84](#))

1.00 ([1.00](#))

Detected mutations

Excluded from this pdf version due to file size limitations.

CFSAN/OAO
BIostatistics AND Bioinformatics Staff

WASTEWATER SARS-COV2 ANALYSIS REPORT

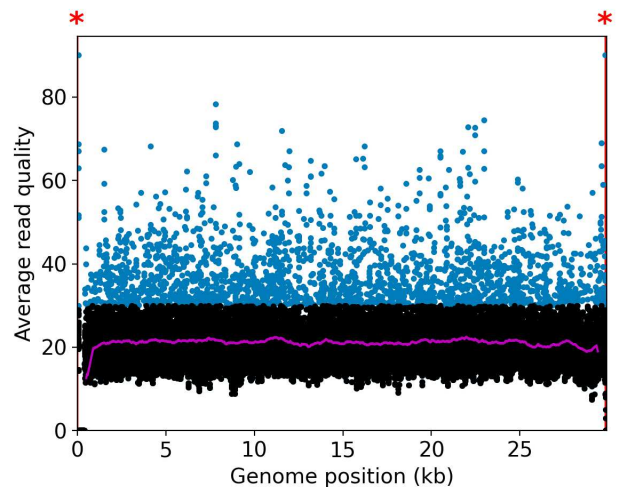
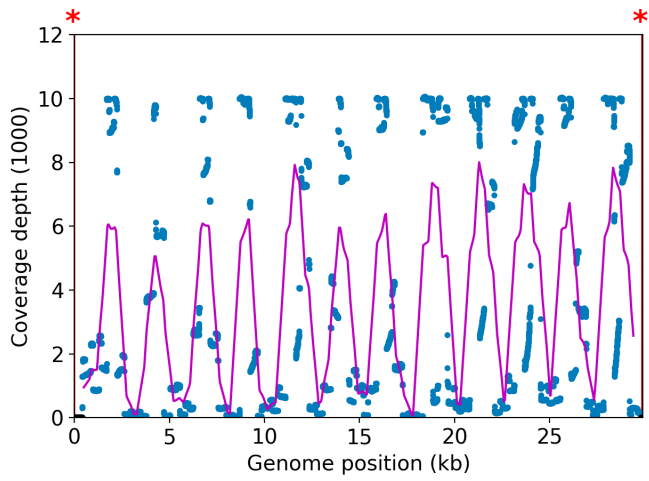
Sample name:	CFSANSMP000119229
Date generated:	2023-02-02, 09:06:51 EST
Timestamp of C-WAP version used:	Tue Jan 31 11:49:22 2023 -0500
Executed by:	Jasmine Amirzadegan (Jasmine.Amirzadegan@fda.hhs.gov)
Executed on:	172.20.44.123 (aka n123.raven.cfsan)

Sequencing summary

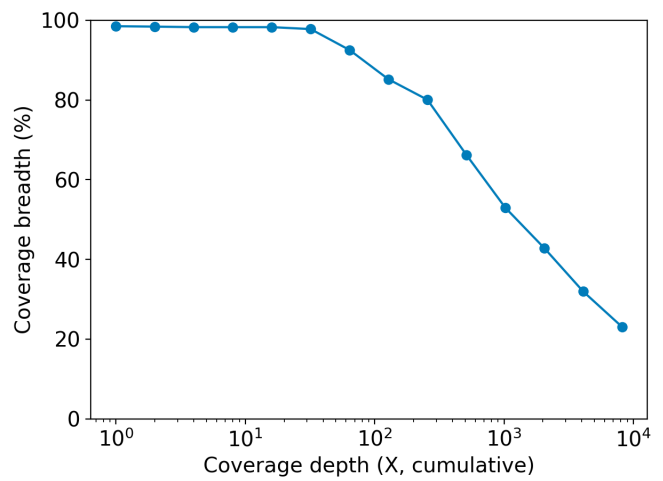
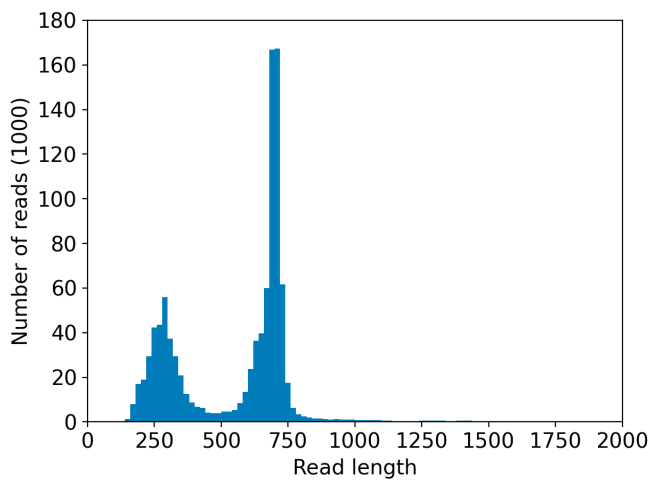
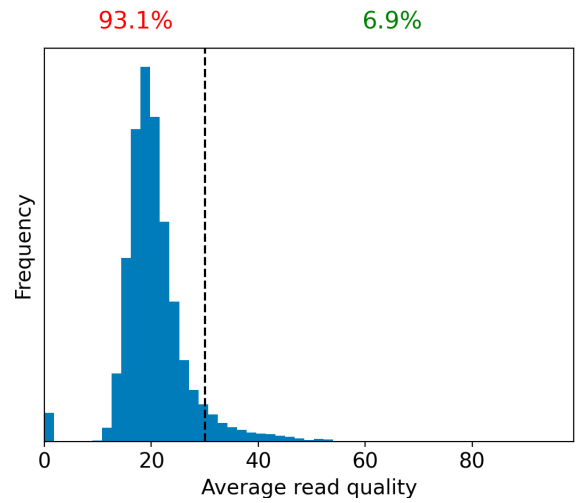
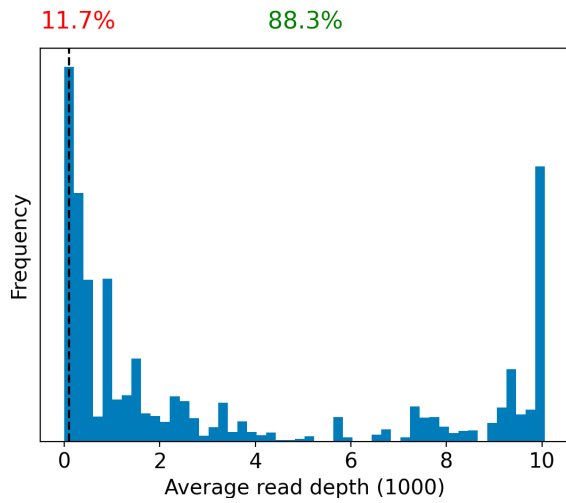
Sequencing chemistry:	Missing with Missing
Source site:	Missing (?.?)
Sampling date:	Missing
Collected by:	Missing
Sequenced by:	Missing
Total number of reads:	1172436
Reads aligned:	794591 (67%)
Average read quality:	19.7
Average read length:	676
Reads passing filter:	794585 (67%)
Average read quality passing filter:	19.7
Average read length passing filter:	676
Average coverage passing filter:	17962X

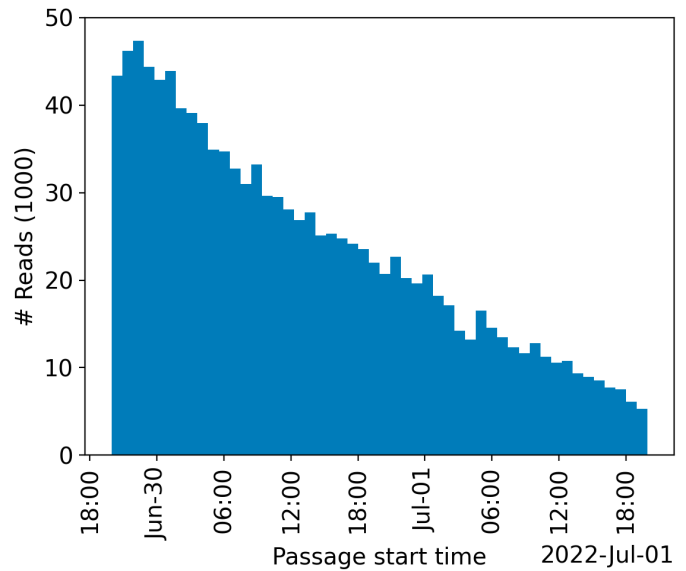
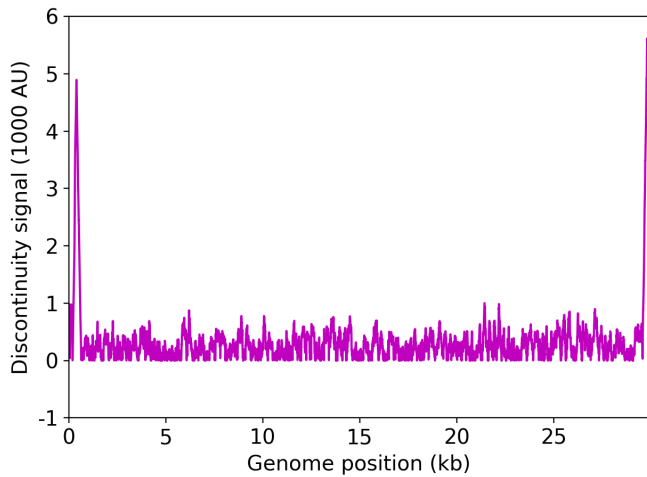
A read passes filter if the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4.

Overall sequence characteristics



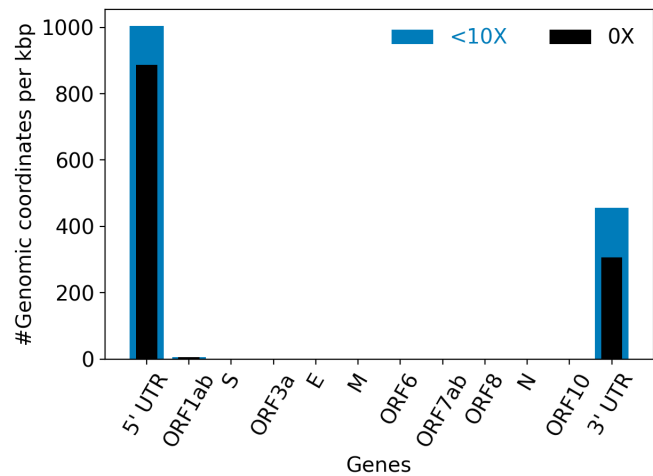
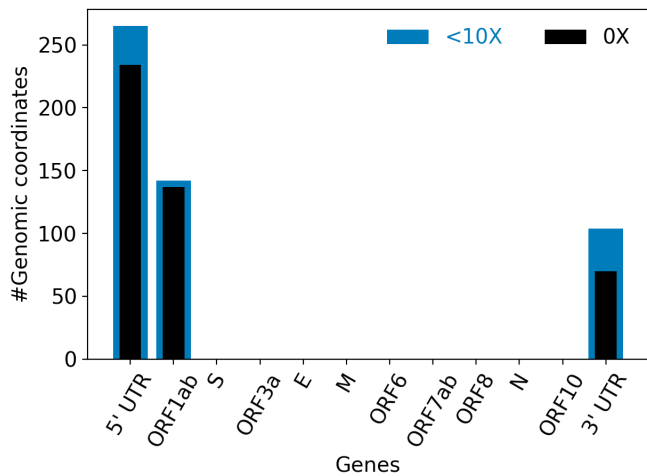
NOTE: The red shaded areas marked with a (*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.





	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	152nt (0%)	152nt (0%)
All genomic coordinates:	441nt (1%)	511nt (1%)
Common SNPs:	1nt (2%)	1nt (2%)
Diverse SNPs:	29nt (12%)	29nt (12%)
Rare SNPs:	60nt (6%)	66nt (7%)

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.



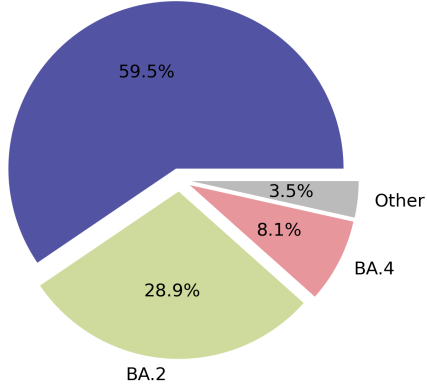
Hits to SARS-Cov2 genome (kraken2):	788012 reads (67.21%)
Hits to human genome (kraken2):	13118 reads (1.12%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
	Coronaviridae (67.21%)

Most abundant organisms (kraken2, family level):

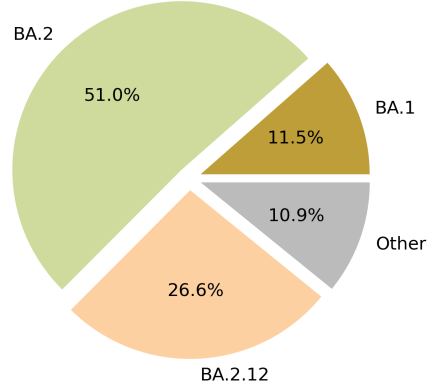
Arcobacteraceae (1.57%)
Hominidae (1.12%)
Enterobacteriaceae (0.40%)
Comamonadaceae (0.30%)

Detected variants (Experimental)

Abundance of variants by linear regression
Omicron



Abundance of variants by kallisto

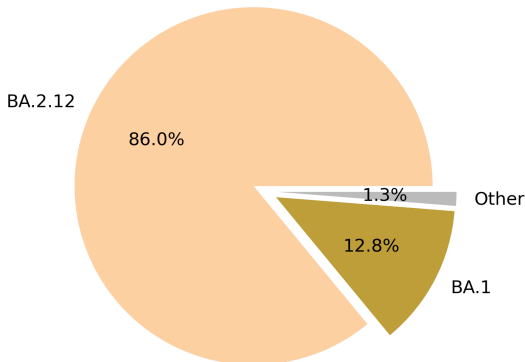


Based on deconvolution, [B.1.1.529](#) is estimated to constitute 59.49% of the viral particles and hence is the most abundant variant in the sample. The R^2 for the linear regression was 0.63. Variants that were detected less than 5% were grouped under "Other"

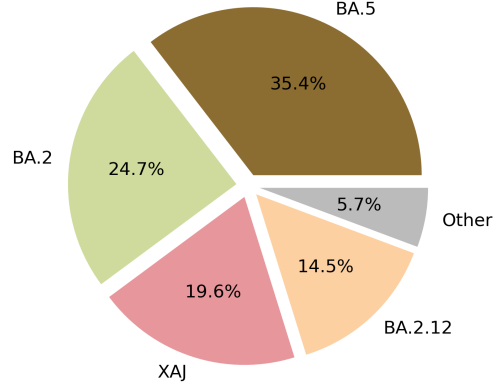
Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [BA.2](#) lineage. If this is a sample consisting of a single source of pathogens or an overwhelming majority of the different sources are infected with the same variant, the sample is dominated by this variant.

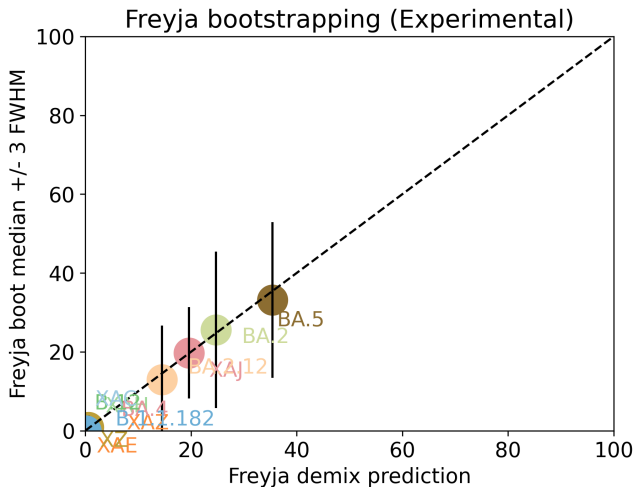
Based on mapping individual reads to the variant consensus sequences in the reference database, kallisto predicts that the sample is dominated by [BA.2](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.

Abundance of variants by kraken2+bracken, using majorCovid DB



Abundance of variants by Freyja





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	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
Characteristic mutations detected	(3 of 13) S:G142D S:L452R S:T478K	(5 of 26) M:D3G NUC:C25000T NUC:C25584T ORF1AB:I3758V S:Q493R	(23 of 31) N:S413R NUC:A20055G NUC:A9424G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:L3027F ORF1AB:L3201F ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:Q493R S:S371F S:T19I S:T376A	(11 of 21) N:S413R NUC:C12880T NUC:C15714T NUC:C26858T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I S:D405N S:Q493R S:S371F	(22 of 31) N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A NUC:G27788T ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G	(20 of 28) N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G

[Jaccard Index](#) is a measure of similarity between two sets A and B, reaching the maximum value of 1 if $A=B$ and minimum value of 0 if $A \cap B = \{\}$. In the c(d) representation below, c represents the Jaccard index of the set of mutations that were experimentally detected for this sample as listed above, whereas d refers to the ideal value of the Jaccard index expected from complete genome coverage without any sequencing errors.

	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
B.1.617.2	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.04 (0.02)	0.05 (0.03)
BA.1	0.00 (0.00)	1.00 (1.00)	0.12 (0.10)	0.07 (0.21)	0.08 (0.08)	0.09 (0.08)
BA.2	0.00 (0.00)	0.12 (0.10)	1.00 (1.00)	0.48 (0.33)	0.67 (0.63)	0.65 (0.59)
BA.3	0.00 (0.00)	0.07 (0.21)	0.48 (0.33)	1.00 (1.00)	0.43 (0.30)	0.41 (0.29)
BA.4	0.04 (0.02)	0.08 (0.08)	0.67 (0.63)	0.43 (0.30)	1.00 (1.00)	0.91 (0.84)

BA.5

0.05 ([0.03](#))

0.09 ([0.08](#))

0.65 ([0.59](#))

0.41 ([0.29](#))

0.91 ([0.84](#))

1.00 ([1.00](#))

Detected mutations

Excluded from this pdf version due to file size limitations.

CFSAN/OAO
BIostatistics AND Bioinformatics Staff

WASTEWATER SARS-COV2 ANALYSIS REPORT

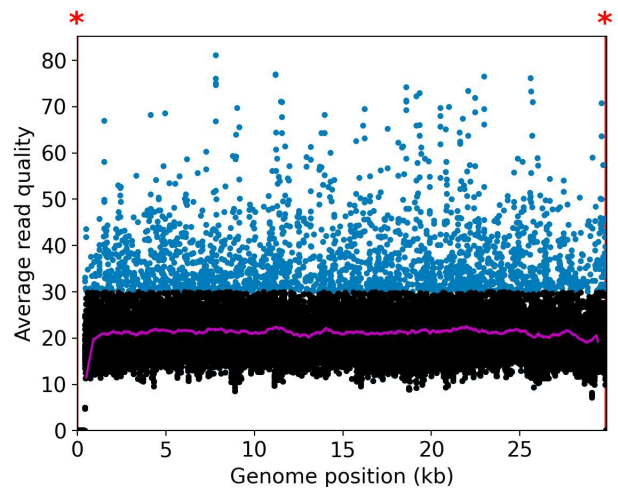
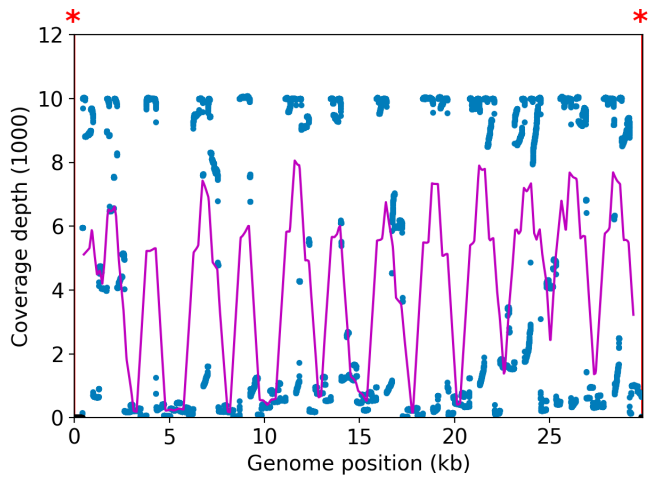
Sample name:	CFSANSMP000119231
Date generated:	2023-02-02, 09:10:35 EST
Timestamp of C-WAP version used:	Tue Jan 31 11:49:22 2023 -0500
Executed by:	Jasmine Amirzadegan (Jasmine.Amirzadegan@fda.hhs.gov)
Executed on:	172.20.44.125 (aka n125.raven.cfsan)

Sequencing summary

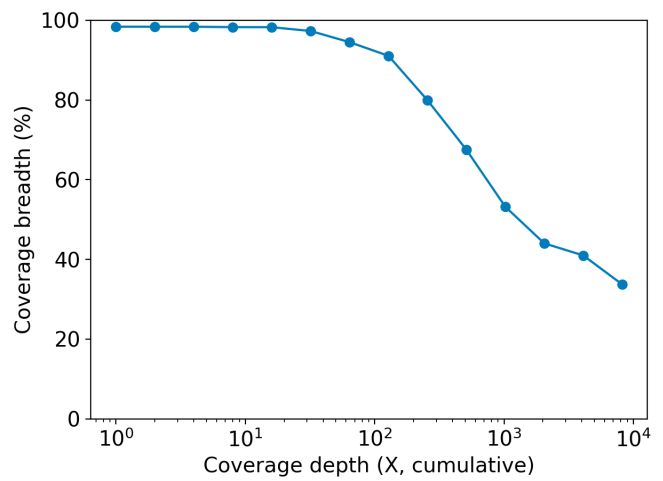
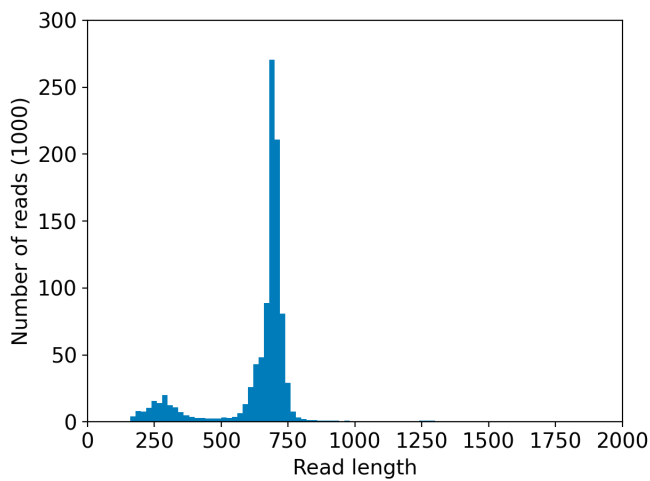
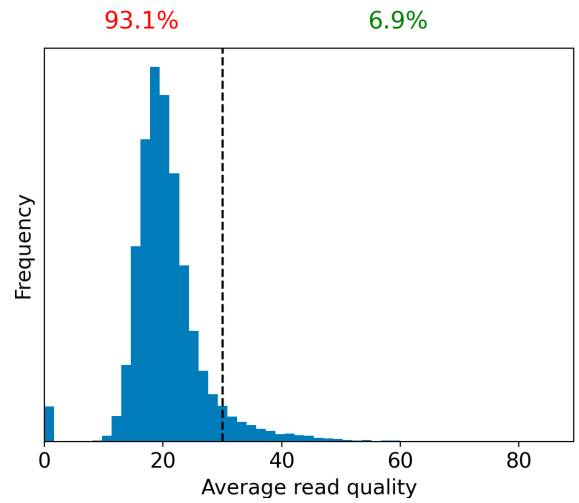
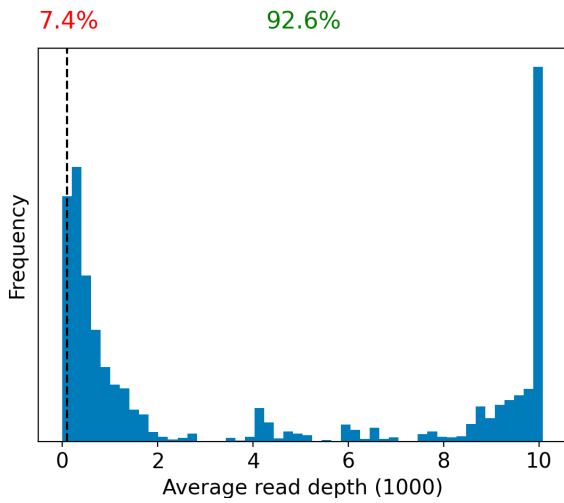
Sequencing chemistry:	Missing with Missing
Source site:	Missing (?.?)
Sampling date:	Missing
Collected by:	Missing
Sequenced by:	Missing
Total number of reads:	1298984
Reads aligned:	1166949 (89%)
Average read quality:	20.0
Average read length:	684
Reads passing filter:	1166947 (89%)
Average read quality passing filter:	20.0
Average read length passing filter:	684
Average coverage passing filter:	26692X

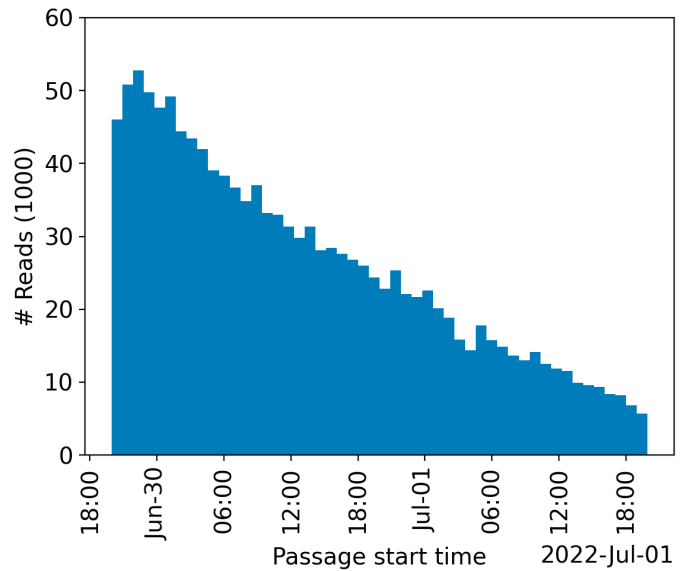
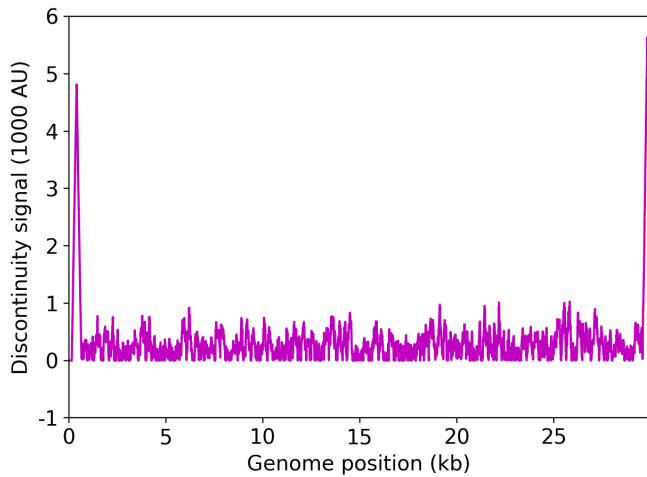
A read passes filter if the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4.

Overall sequence characteristics



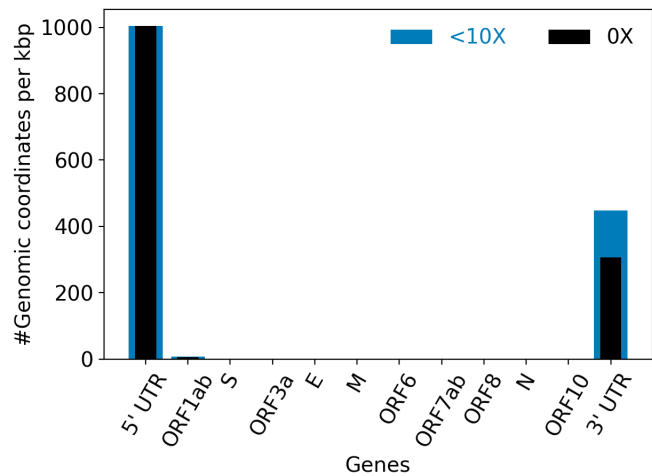
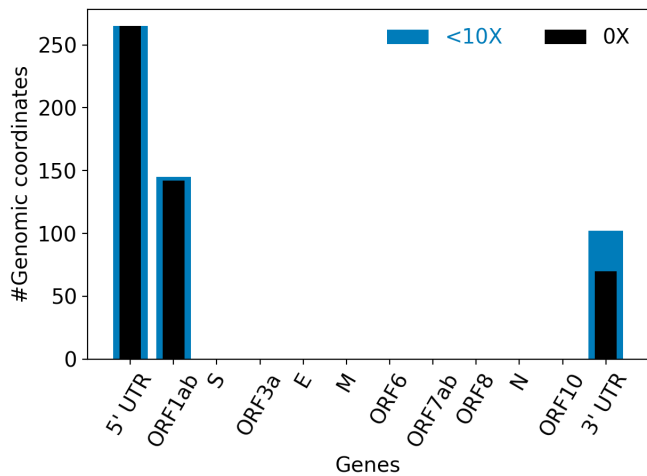
NOTE: The red shaded areas marked with a (*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.





	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	152nt (0%)	152nt (0%)
All genomic coordinates:	477nt (1%)	512nt (1%)
Common SNPs:	1nt (2%)	1nt (2%)
Diverse SNPs:	29nt (12%)	29nt (12%)
Rare SNPs:	60nt (6%)	66nt (7%)

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.



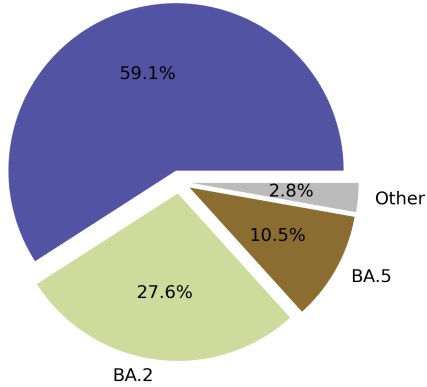
Hits to SARS-Cov2 genome (kraken2):	1157571 reads (89.11%)
Hits to human genome (kraken2):	8174 reads (0.63%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
	Coronaviridae (89.11%)

Most abundant organisms (kraken2, family level):

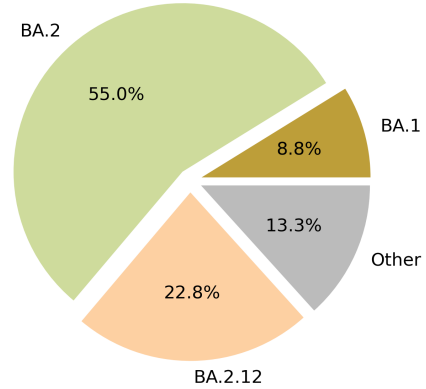
Hominidae (0.63%)
Arcobacteraceae (0.48%)
Enterobacteriaceae (0.12%)
Bacteroidaceae (0.04%)

Detected variants (Experimental)

Abundance of variants by linear regression
Omicron



Abundance of variants by kallisto

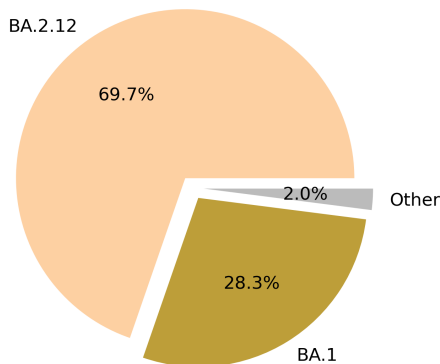


Based on deconvolution, [B.1.1.529](#) is estimated to constitute 59.11% of the viral particles and hence is the most abundant variant in the sample. The R^2 for the linear regression was 0.62. Variants that were detected less than 5% were grouped under "Other"

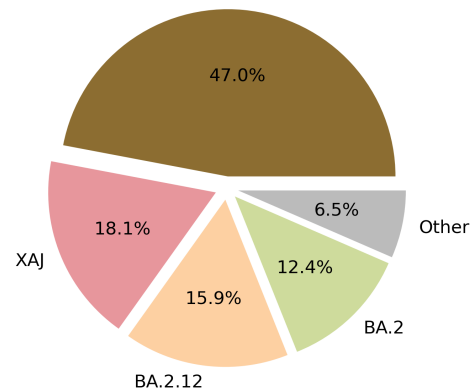
Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [BA.2](#) lineage. If this is a sample consisting of a single source of pathogens or an overwhelming majority of the different sources are infected with the same variant, the sample is dominated by this variant.

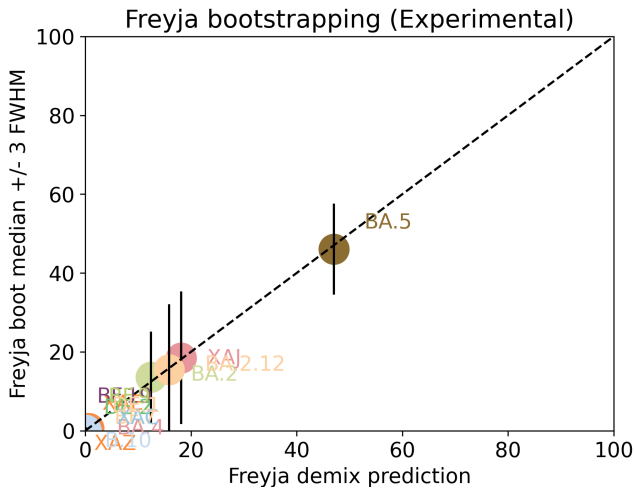
Based on mapping individual reads to the variant consensus sequences in the reference database, kallisto predicts that the sample is dominated by [BA.2](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.

Abundance of variants by kraken2+bracken, using majorCovid DB



Abundance of variants by Freyja
BA.5





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	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
Characteristic mutations detected	(3 of 13) S:G142D S:L452R S:T478K	(5 of 26) M:D3G NUC:C25000T NUC:C25584T ORF1AB:I3758V S:Q493R	(23 of 31) N:S413R NUC:A20055G NUC:A9424G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:L3027F ORF1AB:L3201F ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:Q493R S:R408S S:S371F S:T19I S:T376A	(11 of 21) N:S413R NUC:C12880T NUC:C15714T NUC:C26858T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I S:D405N S:Q493R S:S371F	(22 of 31) N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A NUC:G27788T ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G	(21 of 28) M:D3N N:S413R NUC:A20055G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G

[Jaccard Index](#) is a measure of similarity between two sets A and B, reaching the maximum value of 1 if A=B and minimum value of 0 if $A \cap B = \{\}$. In the c(d) representation below, c represents the Jaccard index of the set of mutations that were experimentally detected for this sample as listed above, whereas d refers to the ideal value of the Jaccard index expected from complete genome coverage without any sequencing errors.

	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
B.1.617.2	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.04 (0.02)	0.04 (0.03)
BA.1	0.00 (0.00)	1.00 (1.00)	0.12 (0.10)	0.07 (0.21)	0.08 (0.08)	0.08 (0.08)
BA.2	0.00 (0.00)	0.12 (0.10)	1.00 (1.00)	0.48 (0.33)	0.67 (0.63)	0.63 (0.59)
BA.3	0.00 (0.00)	0.07 (0.21)	0.48 (0.33)	1.00 (1.00)	0.43 (0.30)	0.39 (0.29)
BA.4	0.04 (0.02)	0.08 (0.08)	0.67 (0.63)	0.43 (0.30)	1.00 (1.00)	0.87 (0.84)

BA.5

0.04 ([0.03](#))

0.08 ([0.08](#))

0.63 ([0.59](#))

0.39 ([0.29](#))

0.87 ([0.84](#))

1.00 ([1.00](#))

Detected mutations

Excluded from this pdf version due to file size limitations.

CFSAN/OAO
BIostatistics AND BIOinformatics STAFF

WASTEWATER SARS-COV2 ANALYSIS REPORT

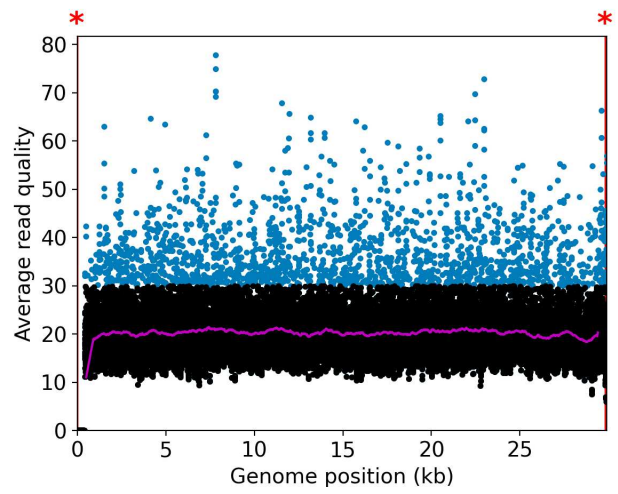
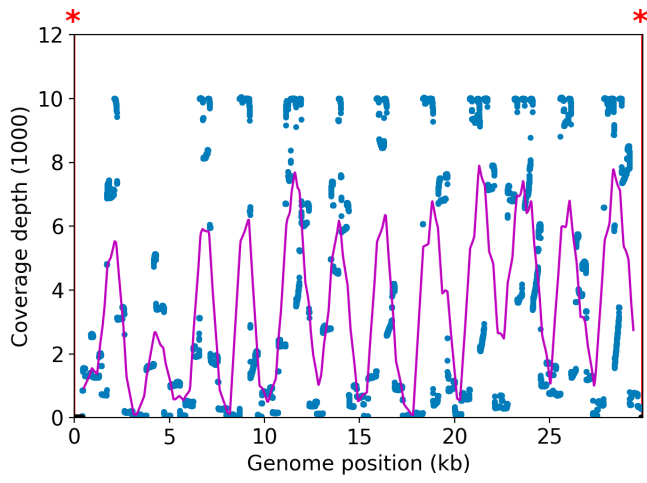
Sample name:	unclassified
Date generated:	2023-02-02, 09:12:00 EST
Timestamp of C-WAP version used:	Tue Jan 31 11:49:22 2023 -0500
Executed by:	Jasmine Amirzadegan (Jasmine.Amirzadegan@fda.hhs.gov)
Executed on:	172.20.44.125 (aka n125.raven.cfsan)

Sequencing summary

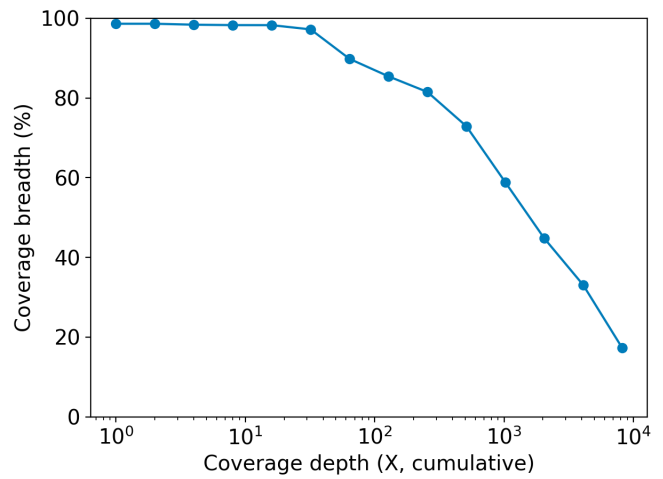
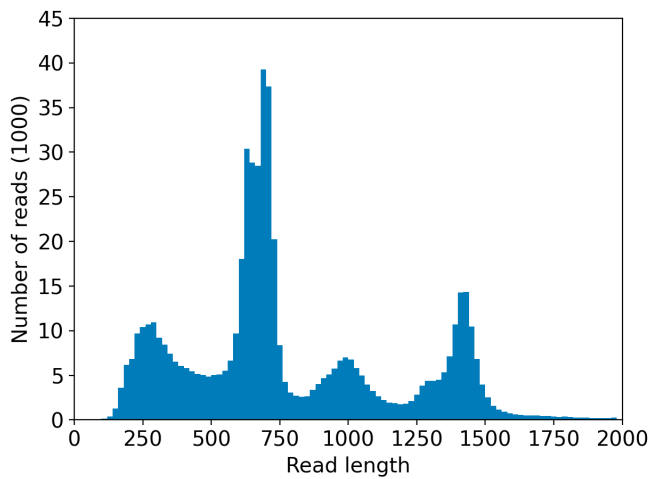
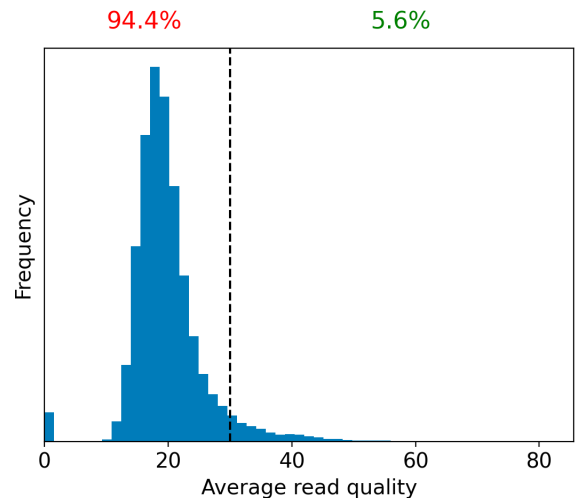
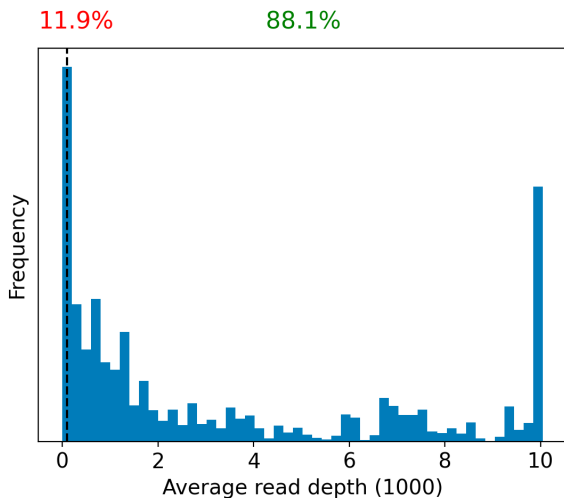
Sequencing chemistry:	Missing with Missing
Source site:	Missing (?.?)
Sampling date:	Missing
Collected by:	Missing
Sequenced by:	Missing
Total number of reads:	569329
Reads aligned:	462289 (81%)
Average read quality:	19.3
Average read length:	883
Reads passing filter:	462282 (81%)
Average read quality passing filter:	19.3
Average read length passing filter:	883
Average coverage passing filter:	13650X

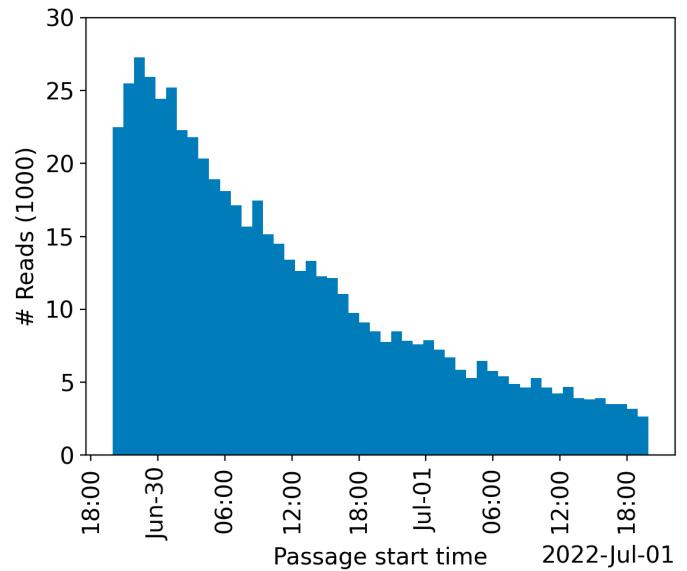
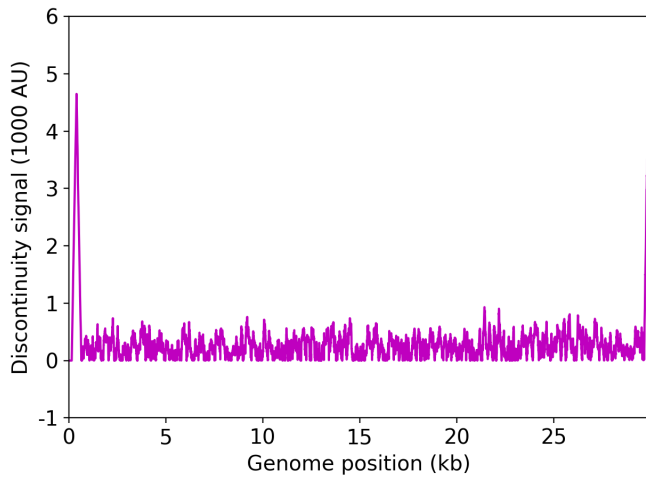
A read passes filter if the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4.

Overall sequence characteristics



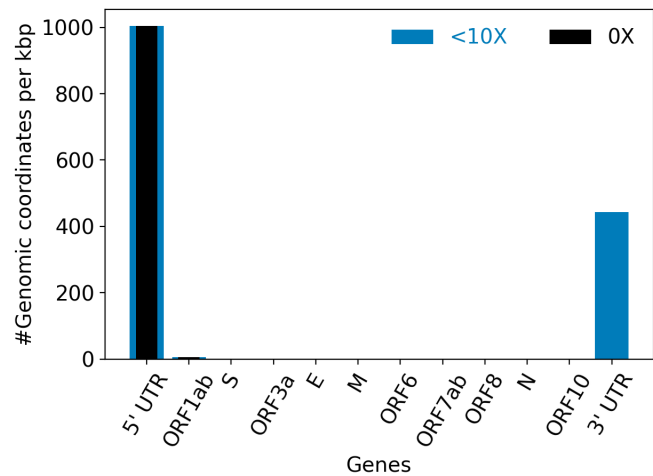
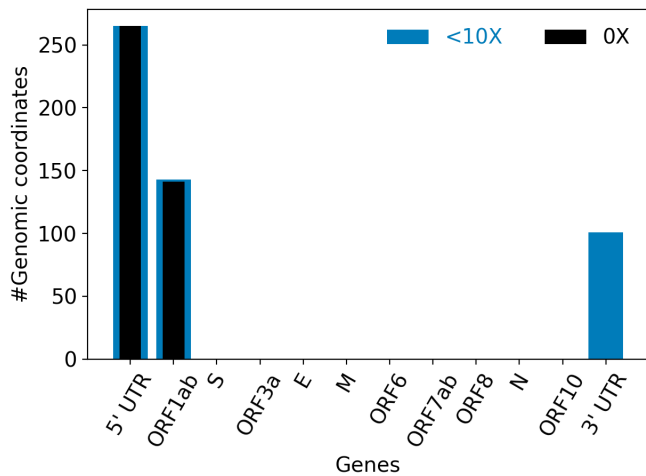
NOTE: The red shaded areas marked with a (*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.





	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	152nt (0%)	152nt (0%)
All genomic coordinates:	406nt (1%)	509nt (1%)
Common SNPs:	1nt (2%)	1nt (2%)
Diverse SNPs:	2nt (0%)	29nt (12%)
Rare SNPs:	49nt (5%)	66nt (7%)

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.

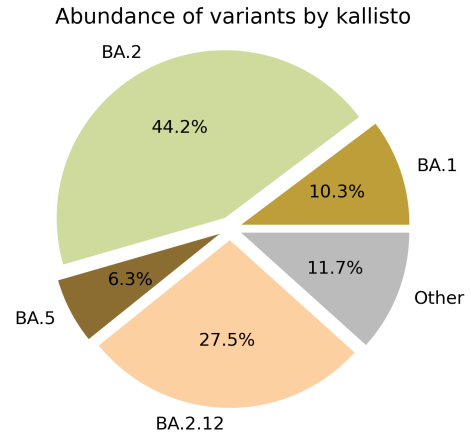
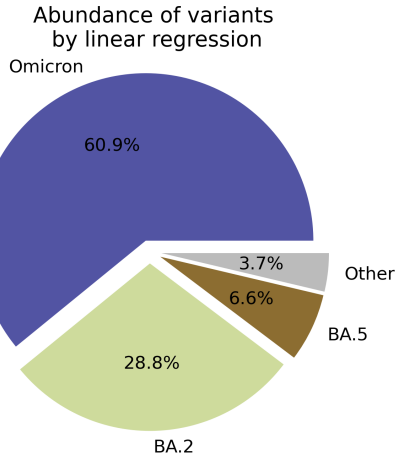


Hits to SARS-Cov2 genome (kraken2):	436270 reads (76.63%)
Hits to human genome (kraken2):	45590 reads (8.01%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
	Coronaviridae (76.63%)

Most abundant organisms (kraken2, family level):

Hominidae (8.01%)
Arcobacteraceae (1.91%)
Enterobacteriaceae (0.59%)
Bacteroidaceae (0.18%)

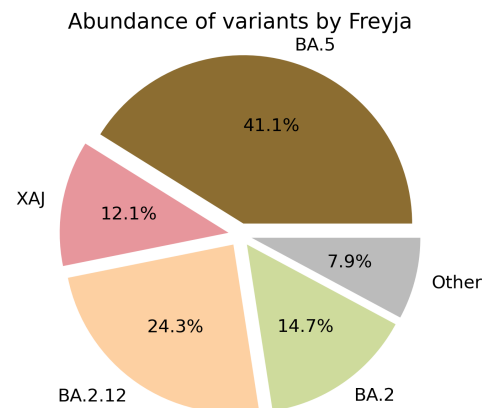
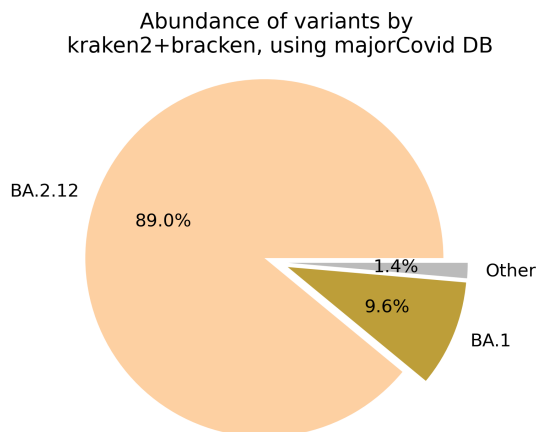
Detected variants (Experimental)

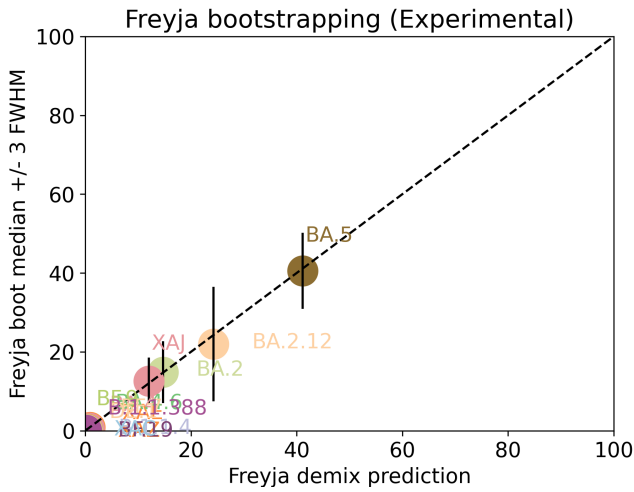


Based on deconvolution, [B.1.1.529](#) is estimated to constitute 60.90% of the viral particles and hence is the most abundant variant in the sample. The R^2 for the linear regression was 0.63. Variants that were detected less than 5% were grouped under "Other"

Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [BA.2](#) 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.2](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.





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	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
Characteristic mutations detected	(3 of 13) S:G142D S:L452R S:T478K	(5 of 26) M:D3G NUC:C25000T NUC:C25584T ORF1AB:I3758V S:Q493R	(22 of 31) N:S413R NUC:A9424G NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:L3027F ORF1AB:L3201F ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:Q493R S:R408S S:S371F S:T19I S:T376A	(11 of 21) N:S413R NUC:C12880T NUC:C15714T NUC:C26858T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I S:D405N S:Q493R S:S371F	(21 of 31) N:S413R NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C26858T NUC:C4321T NUC:G10447A NUC:G27788T ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G	(19 of 28) N:S413R NUC:C10198T NUC:C12880T NUC:C15714T NUC:C25000T NUC:C25584T NUC:C4321T NUC:G10447A ORF1AB:G1307S ORF1AB:S135R ORF1AB:T3090I ORF1AB:T842I S:D405N S:F486V S:L452R S:S371F S:T19I S:T376A S:V213G

Jaccard Index is a measure of similarity between two sets A and B, reaching the maximum value of 1 if A=B and minimum value of 0 if $A \cap B = \{\}$. In the c(d) representation below, c represents the Jaccard index of the set of mutations that were experimentally detected for this sample as listed above, whereas d refers to the ideal value of the Jaccard index expected from complete genome coverage without any sequencing errors.

	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
B.1.617.2	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.04 (0.02)	0.05 (0.03)
BA.1	0.00 (0.00)	1.00 (1.00)	0.12 (0.10)	0.07 (0.21)	0.08 (0.08)	0.09 (0.08)
BA.2	0.00 (0.00)	0.12 (0.10)	1.00 (1.00)	0.50 (0.33)	0.65 (0.63)	0.64 (0.59)
BA.3	0.00 (0.00)	0.07 (0.21)	0.50 (0.33)	1.00 (1.00)	0.45 (0.30)	0.43 (0.29)
BA.4	0.04 (0.02)	0.08 (0.08)	0.65 (0.63)	0.45 (0.30)	1.00 (1.00)	0.90 (0.84)
BA.5	0.05 (0.03)	0.09 (0.08)	0.64 (0.59)	0.43 (0.29)	0.90 (0.84)	1.00 (1.00)

BA.5

0.05 ([0.03](#))

0.09 ([0.08](#))

0.64 ([0.59](#))

0.43 ([0.29](#))

0.90 ([0.84](#))

1.00 ([1.00](#))

Detected mutations

Excluded from this pdf version due to file size limitations.

CFSAN/OAO
BIostatistics AND BIOinformatics STAFF

WASTEWATER SARS-COV2 ANALYSIS REPORT

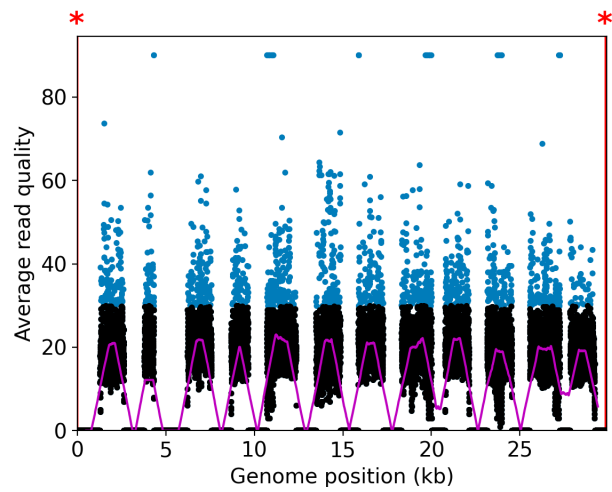
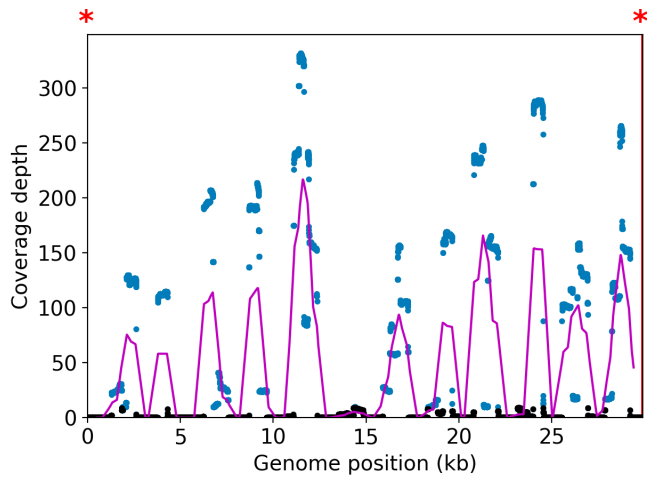
Sample name:	Water
Date generated:	2023-02-02, 08:50:56 EST
Timestamp of C-WAP version used:	Tue Jan 31 11:49:22 2023 -0500
Executed by:	Jasmine Amirzadegan (Jasmine.Amirzadegan@fda.hhs.gov)
Executed on:	172.20.44.121 (aka n121.raven.cfsan)

Sequencing summary

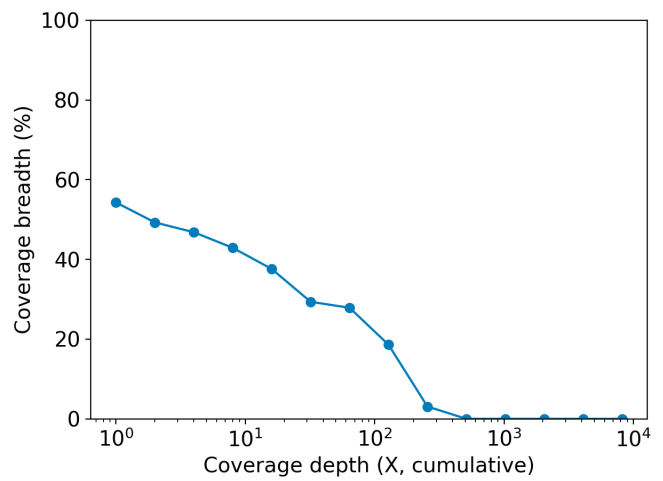
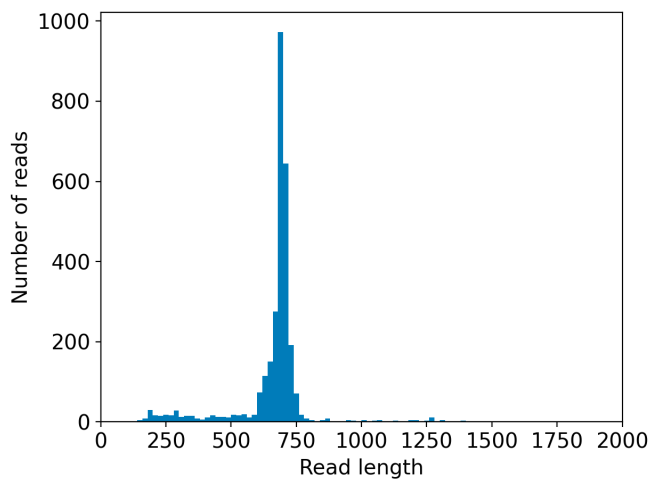
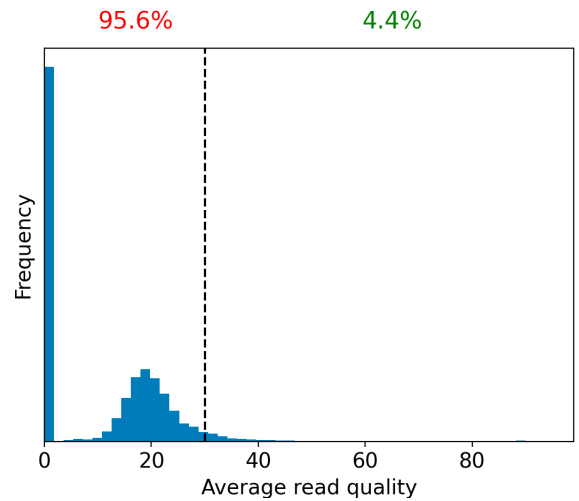
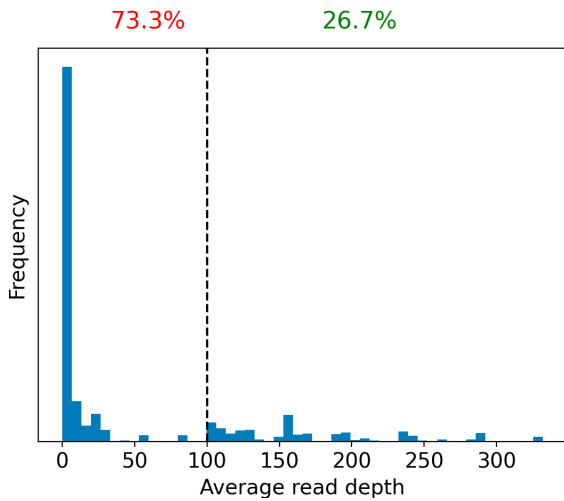
Sequencing chemistry:	Missing with Missing
Source site:	Missing (?.?)
Sampling date:	Missing
Collected by:	Missing
Sequenced by:	Missing
Total number of reads:	2964
Reads aligned:	2896 (97%)
Average read quality:	20.8
Average read length:	688
Reads passing filter:	2896 (97%)
Average read quality passing filter:	20.8
Average read length passing filter:	688
Average coverage passing filter:	66X

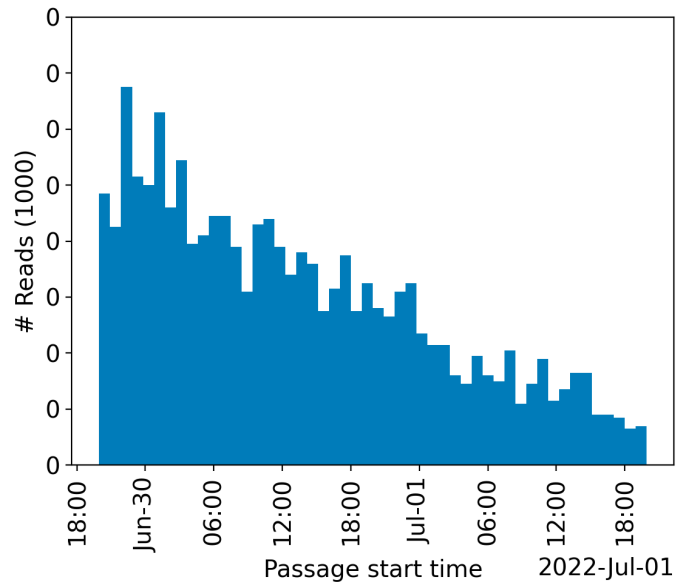
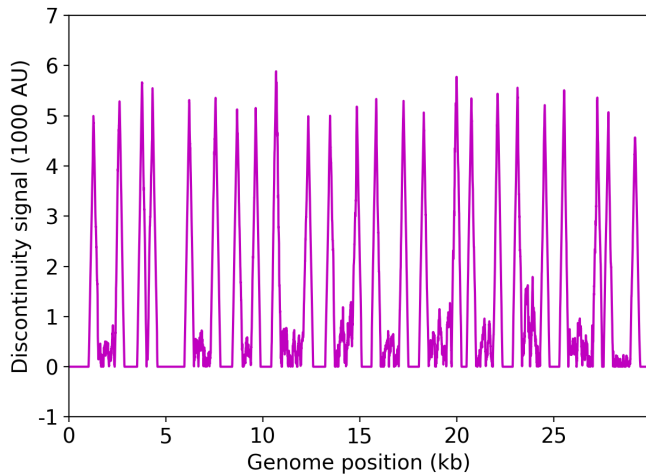
A read passes filter if the read length after adaptor trimming ≥ 30 and minimum read quality ≥ 20 within a sliding window of width 4.

Overall sequence characteristics



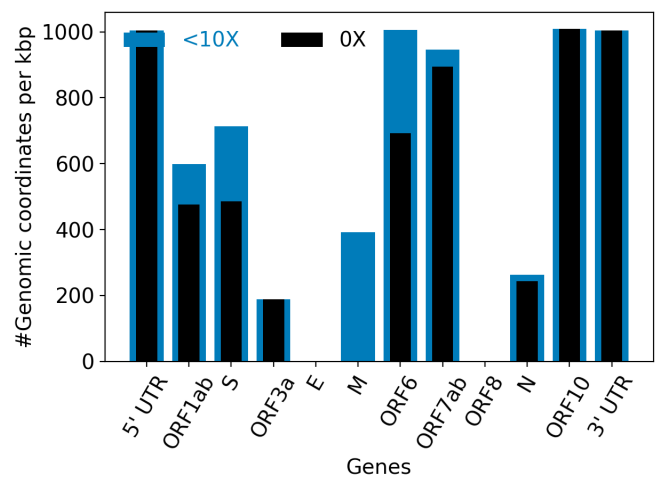
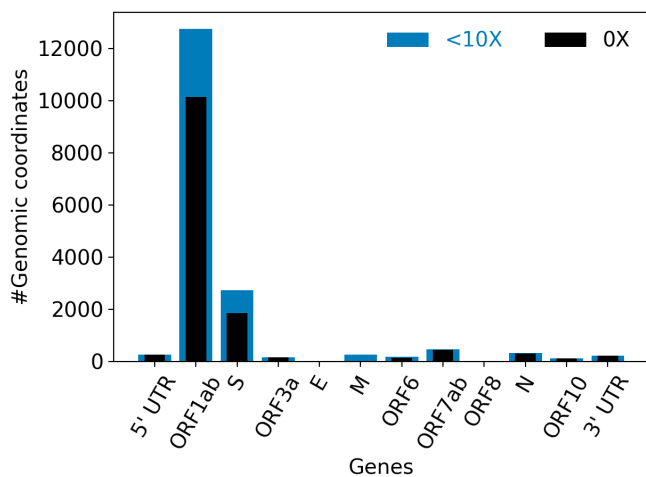
NOTE: The red shaded areas marked with a (*) are not covered by the design of the library preparation kit and hence excluded from analyses. Magenta curves represent moving average with a window width of 1kb.





	Uncovered coordinates (0X)	Poorly covered coordinates (<10X)
# Inaccessible genomic coordinates by kit design:	152nt (0%)	152nt (0%)
All genomic coordinates:	13665nt (45%)	17534nt (58%)
Common SNPs:	16nt (44%)	22nt (61%)
Diverse SNPs:	91nt (40%)	187nt (83%)
Rare SNPs:	681nt (73%)	807nt (87%)

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.

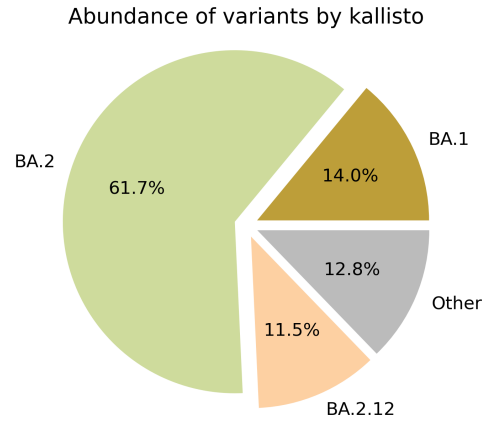
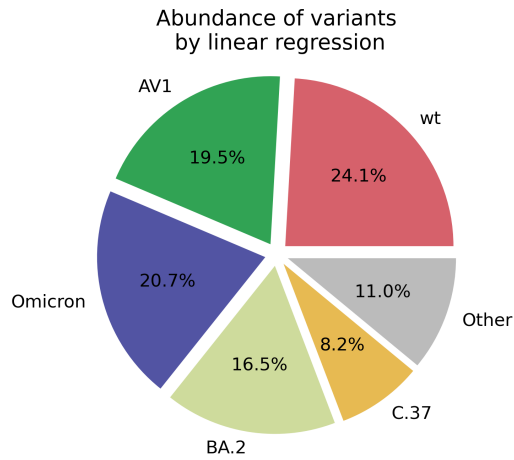


Hits to SARS-Cov2 genome (kraken2):	2866 reads (96.69%)
Hits to human genome (kraken2):	11 reads (0.37%)
Hits to synthetic sequences (kraken2, taxid 28384):	0 reads (0.00%)
	Coronaviridae (96.69%)

Most abundant organisms (kraken2, family level):

Hominidae (0.37%)
Arcobacteraceae (0.03%)
Oscillospiraceae (0.03%)
Halorubraceae (0.03%)

Detected variants (Experimental)

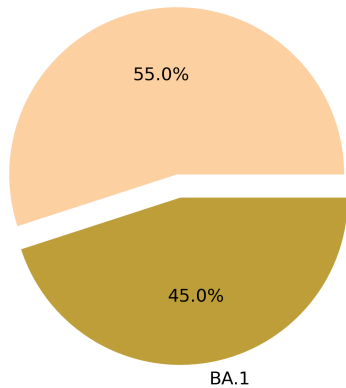


Based on deconvolution, [wt](#) is estimated to constitute 23.76% of the viral particles and hence is the most abundant variant in the sample. The R^2 for the linear regression was 0.19. Variants that were detected less than 5% were grouped under "Other"

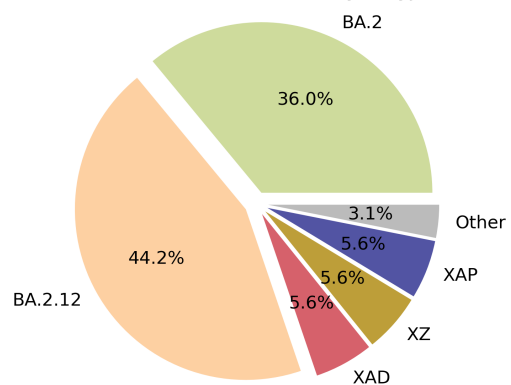
Based on the consensus sequence of the observed reads, the "ensemble-averaged sequence" most closely resembles the [B.1.1](#) lineage. If this is a sample consisting of a single source of pathogens or an overwhelming majority of the different sources are infected with the same variant, the sample is dominated by this variant.

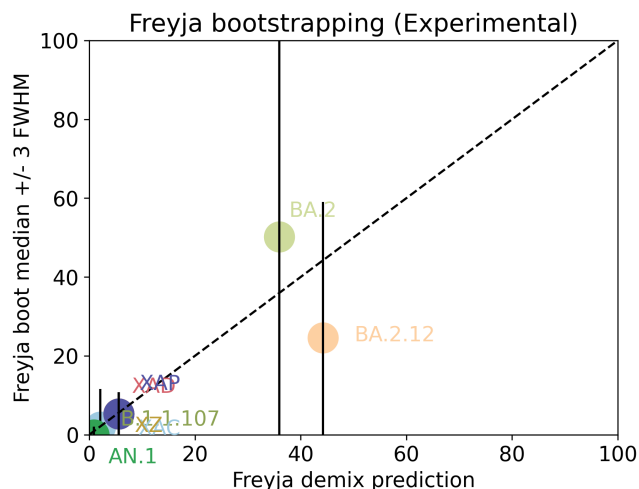
Based on mapping individual reads to the variant consensus sequences in the reference database, kallisto predicts that the sample is dominated by [BA.2](#) lineage. Accuracy of this measure is expected to improve if the input data consists of long reads as opposed to convolution.

Abundance of variants by kraken2+bracken, using majorCovid DB
BA.2.12



Abundance of variants by Freyja





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	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
Characteristic mutations detected	(1 of 13) S:G142D	(3 of 26) M:D3G NUC:C25584T ORF1AB:I3758V	(6 of 31) NUC:C25584T NUC:C26858T ORF1AB:G1307S ORF1AB:L3027F ORF1AB:T3090I S:T19I	(3 of 21) NUC:C26858T ORF1AB:G1307S ORF1AB:T3090I	(5 of 31) NUC:C25584T NUC:C26858T ORF1AB:G1307S ORF1AB:T3090I S:T19I	(4 of 28) NUC:C25584T ORF1AB:G1307S ORF1AB:T3090I S:T19I

[Jaccard Index](#) is a measure of similarity between two sets A and B, reaching the maximum value of 1 if $A=B$ and minimum value of 0 if $A \cap B = \{\}$. In the c(d) representation below, c represents the Jaccard index of the set of mutations that were experimentally detected for this sample as listed above, whereas d refers to the ideal value of the Jaccard index expected from complete genome coverage without any sequencing errors.

	B.1.617.2	BA.1	BA.2	BA.3	BA.4	BA.5
B.1.617.2	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.02)	0.00 (0.03)
BA.1	0.00 (0.00)	1.00 (1.00)	0.12 (0.10)	0.00 (0.21)	0.14 (0.08)	0.17 (0.08)
BA.2	0.00 (0.00)	0.12 (0.10)	1.00 (1.00)	0.50 (0.33)	0.83 (0.63)	0.67 (0.59)
BA.3	0.00 (0.00)	0.00 (0.21)	0.50 (0.33)	1.00 (1.00)	0.60 (0.30)	0.40 (0.29)
BA.4	0.00 (0.02)	0.14 (0.08)	0.83 (0.63)	0.60 (0.30)	1.00 (1.00)	0.80 (0.84)
BA.5	0.00 (0.03)	0.17 (0.08)	0.67 (0.59)	0.40 (0.29)	0.80 (0.84)	1.00 (1.00)

Detected mutations

Excluded from this pdf version due to file size limitations.