

COVID-19 subject AHS19005093

2021-04-17

The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with $> 90\%$ sequence coverage.

Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (≥ 5 reads)
VSP0638	composite	NA	VTM	2021-01-22	22.30	B.1.243	99.8%	99.2%
VSP0638-1	single experiment	NA	VTM	2021-01-22	1.14	NA	12.9%	11.6%
VSP0638-2	single experiment	NA	VTM	2021-01-22	22.30	B.1.243	99.8%	99.2%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/USA-WA1-2020) are shared across subject samples where the percent variance is colored. Variants are called if a variant position is covered by 5 or more reads, the alternative base is found in > 50% of read pairs and the variant yields a PHRED score > 20. Gray tiles denote positions where the variant was not the major variant or no variants were found. The relative base compositions of each experiment used to calculate tiles are shown in the following plot where the total number of position reads are shown atop of each plot.



VTM
2021-01-22

241 intergenic	0	903
913 orf1ab silent	991	1974
2110 orf1ab silent	0	1599
3037 orf1ab silent	8226	832
5986 orf1ab silent	0	838
6954 orf1ab I2230T	0	320
7042 orf1ab M2259I	0	391
8782 orf1ab silent	0	1780
11102 orf1ab P3613S	0	12249
14408 orf1ab P314L	0	1574
18060 orf1ab silent	0	1304
19875 orf1ab silent	0	12473
20268 orf1ab silent	0	4006
23063 S N501Y	0	1196
23403 S D614G	0	12366
23604 S P681H	0	1288
23709 S T716I	0	1202
24076 S silent	0	7290
24914 S D1118H	0	2754
27972 orf8 Q27stop	0	2183
28048 orf8 R52I	0	1745
28144 orf8 S84L	0	9533
28854 N S194L	0	771
29044 N silent	0	1038

Base change

- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

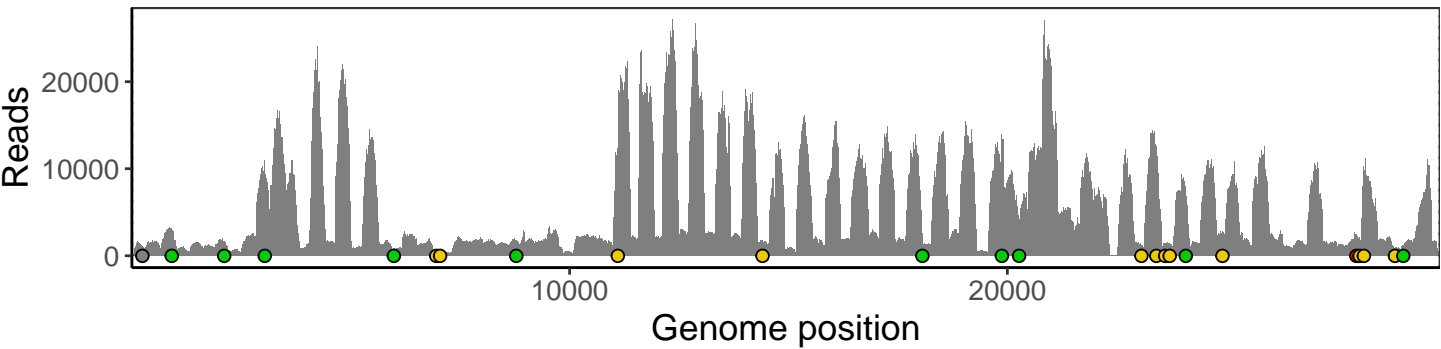
VSP0638-1

VSP0638-2

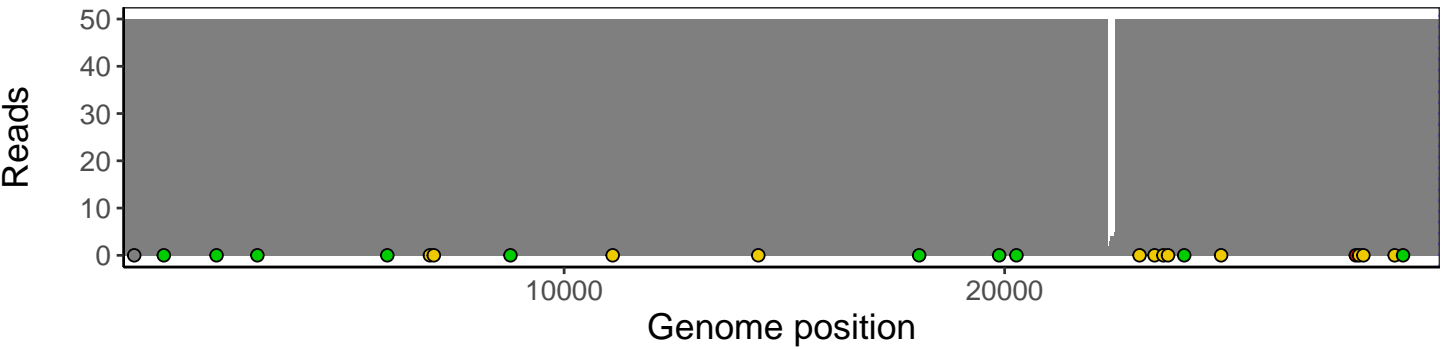
Analyses of individual experiments and composite results

VSP0638 | 2021-01-22 | VTM | H2101191203 | composite result

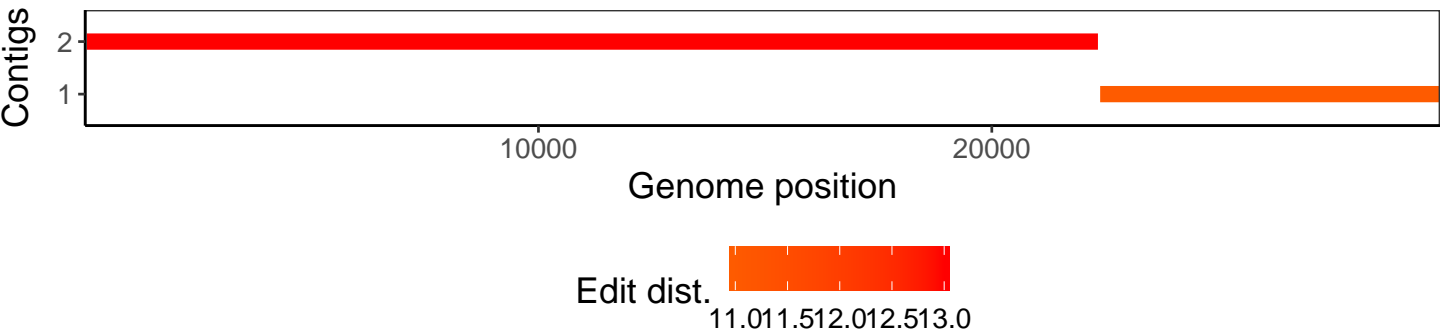
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



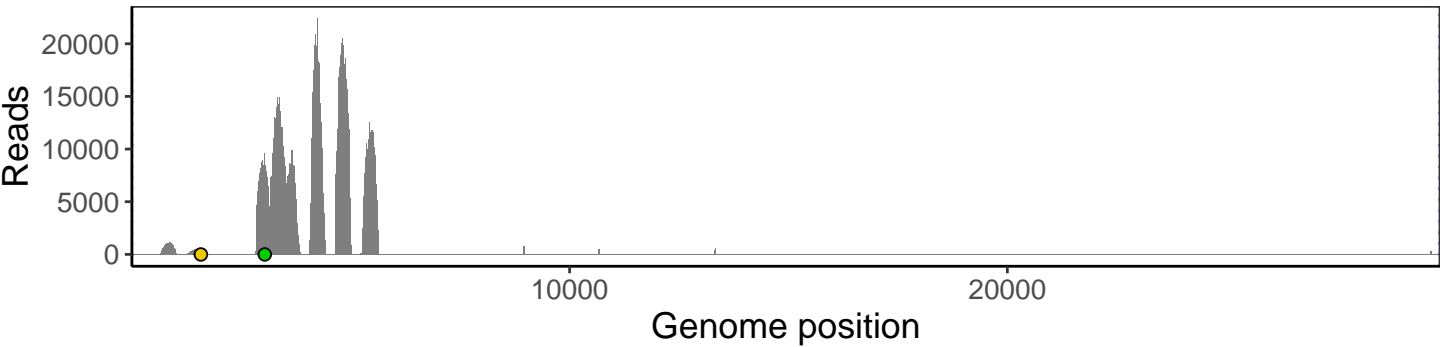
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



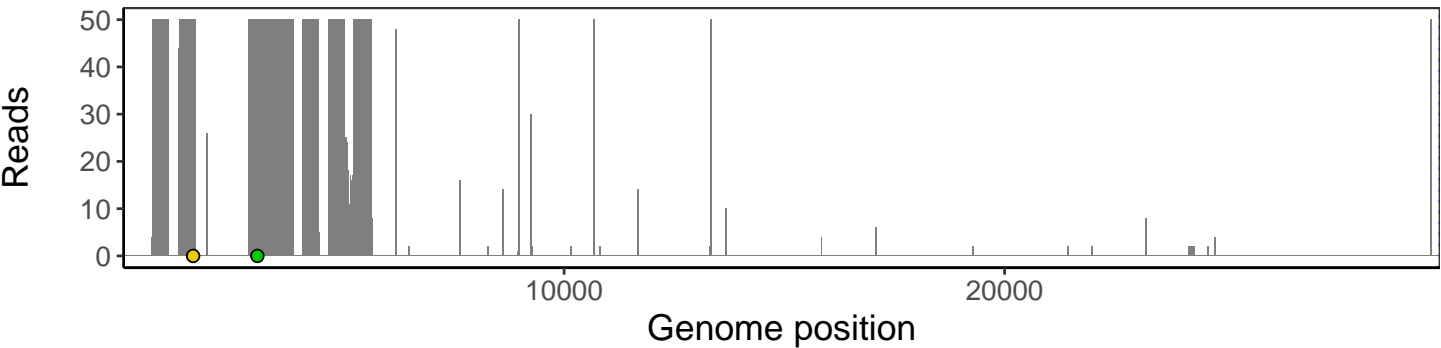
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



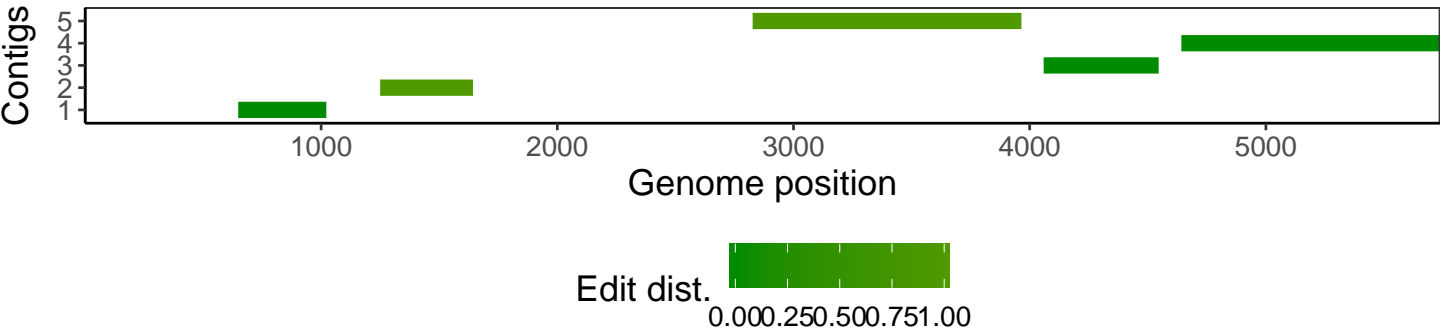
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



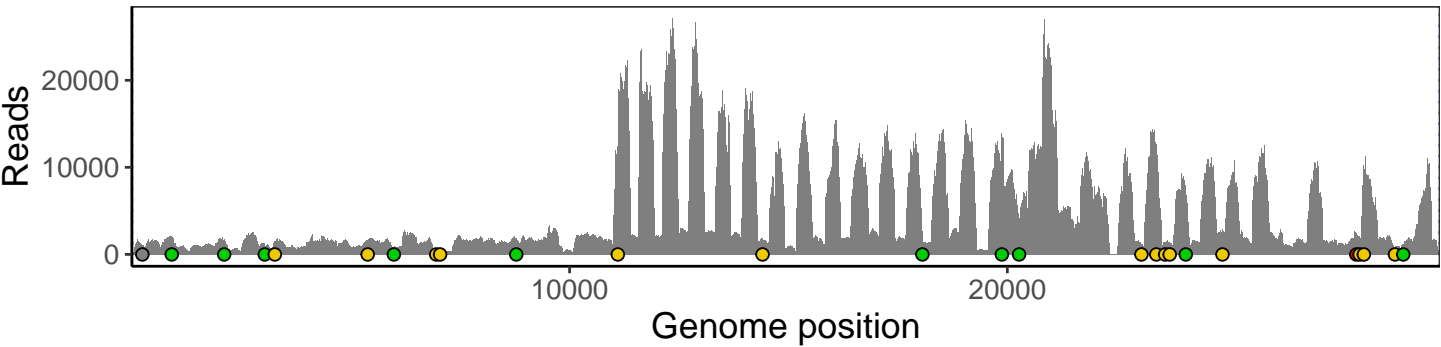
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



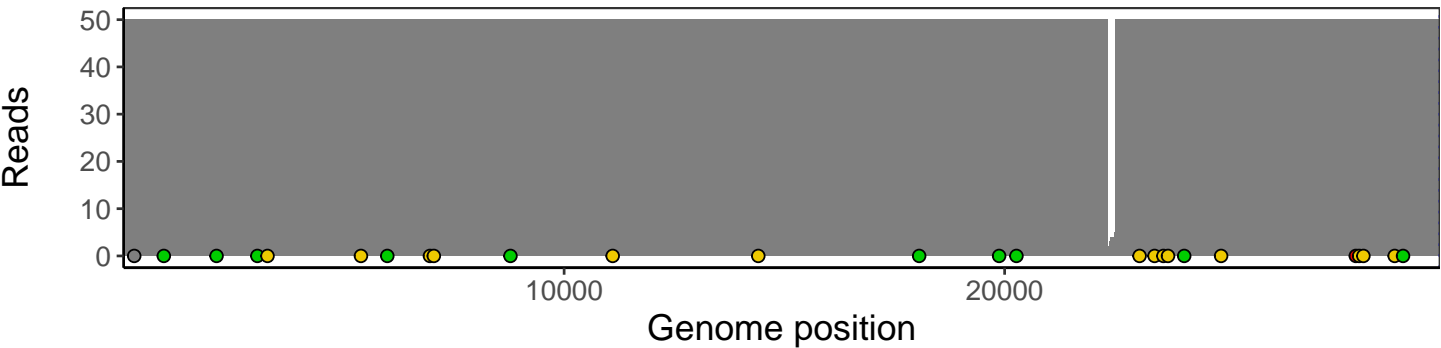
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



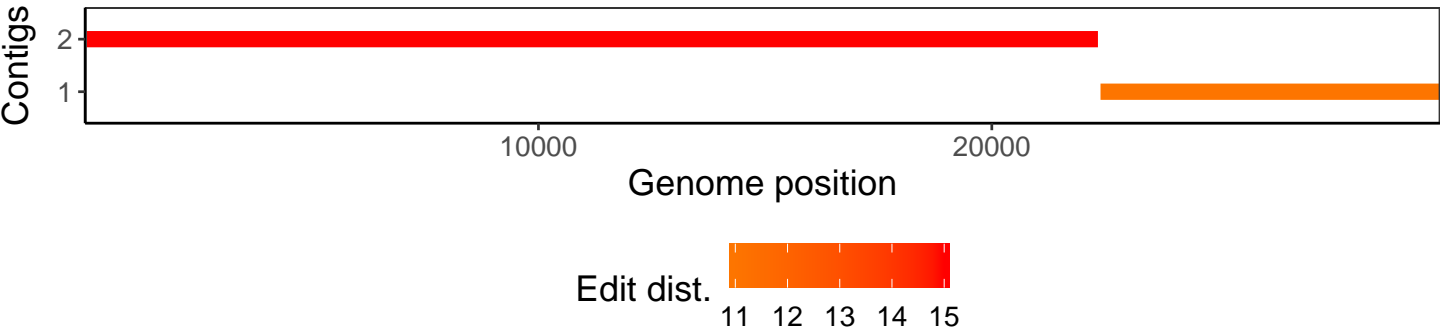
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



Software environment

Software/R package	Version
R	3.4.0
bwa	0.7.17-r1198-dirty
samtools	1.10 Using htlib 1.10
bcftools	1.10.2-34-g1a12af0-dirty Using htlib 1.10.2-57-gf58a6f3
pangolin	2.3.8
genbankr	1.4.0
optparse	1.6.0
forcats	0.3.0
stringr	1.4.0
dplyr	0.8.1
purrr	0.2.5
readr	1.1.1
tidyr	0.8.1
tibble	2.1.2
ggplot2	3.0.0
tidyverse	1.2.1
ShortRead	1.34.2
GenomicAlignments	1.12.2
SummarizedExperiment	1.6.5
DelayedArray	0.2.7
matrixStats	0.54.0
Biobase	2.36.2
Rsamtools	1.28.0
GenomicRanges	1.28.6
GenomeInfoDb	1.12.3
Biostrings	2.44.2
XVector	0.16.0
IRanges	2.10.5
S4Vectors	0.14.7
BiocParallel	1.10.1
BiocGenerics	0.22.1