

COVID-19 subject AHS21001204

2021-03-29

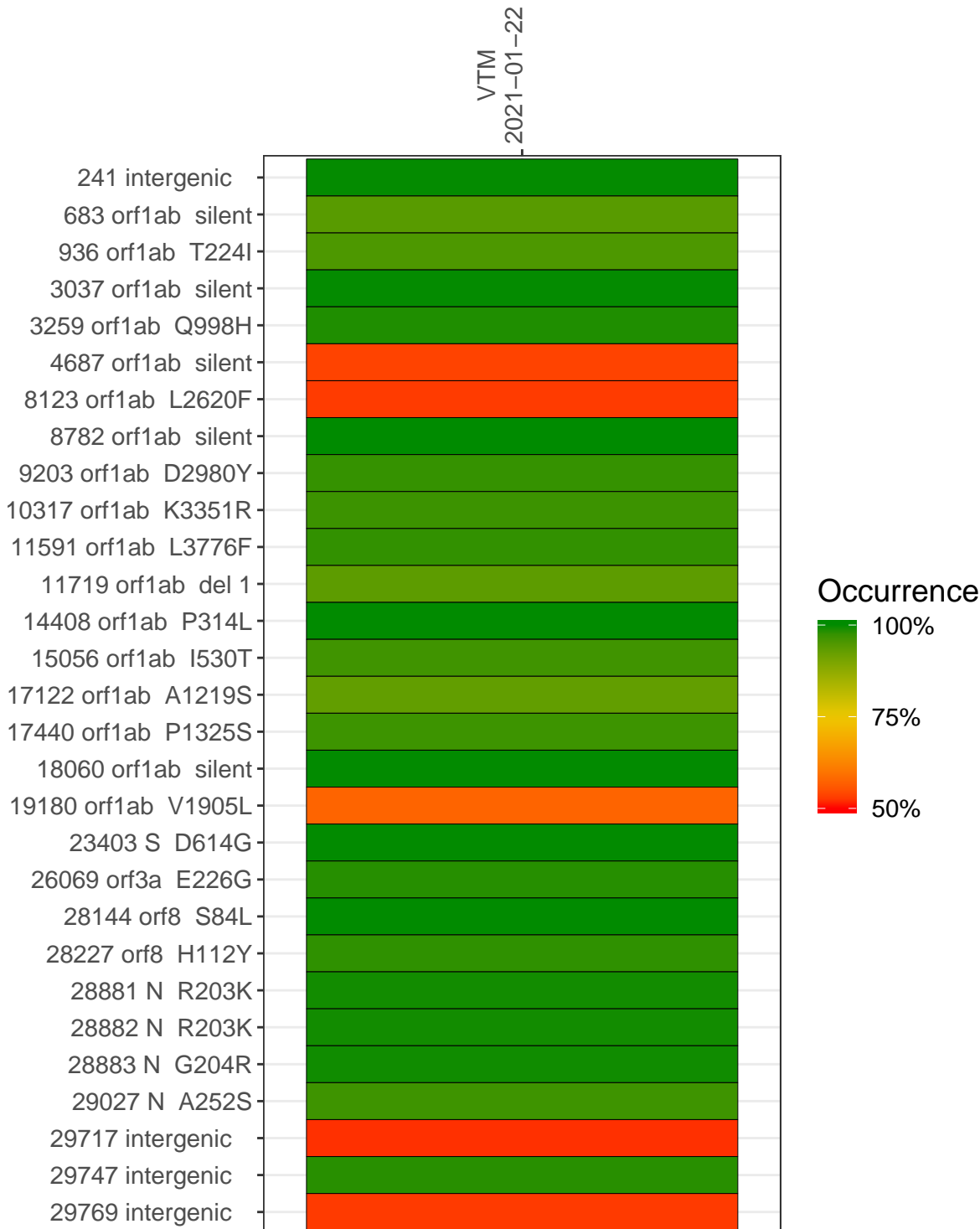
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP0637	composite	NA	VTM	2021-01-22	29.93	B.1.1.304	99.8%	99.8%
VSP0637-1	single experiment	NA	VTM	2021-01-22	1.08	NA	36.4%	34.8%
VSP0637-2	single experiment	NA	VTM	2021-01-22	29.82	B.1.1.304	99.8%	99.8%

Variants shared across samples

The heat map below shows how variants (reference genome 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

241 intergenic	6152	68
683 orf1ab silent	1069	67
936 orf1ab T224I	3028	150
3037 orf1ab silent	1823	101
3259 orf1ab Q998H	3232	6776
4687 orf1ab silent	4	8399
8123 orf1ab L2620F	0	142
8782 orf1ab silent	0	161
9203 orf1ab D2980Y	0	6054
10317 orf1ab K3351R	14	8795
11591 orf1ab L3776F	0	11448
11719 orf1ab del 1	0	4771
14408 orf1ab P314L	7051	208
15056 orf1ab I530T	2689	96
17122 orf1ab A1219S	0	4536
17440 orf1ab P1325S	0	4472
18060 orf1ab silent	0	15625
19180 orf1ab V1905L	0	153
23403 S D614G	3470	198
26069 orf3a E226G	4142	13920
28144 orf8 S84L	693	7545
28227 orf8 H112Y	0	5956
28881 N R203K	2236	1916
28882 N R203K	2231	1908
28883 N G204R	2233	1930
29027 N A252S	572	1236
29717 intergenic	0	10418
29747 intergenic	0	11275
29769 intergenic	0	10481

Base change

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

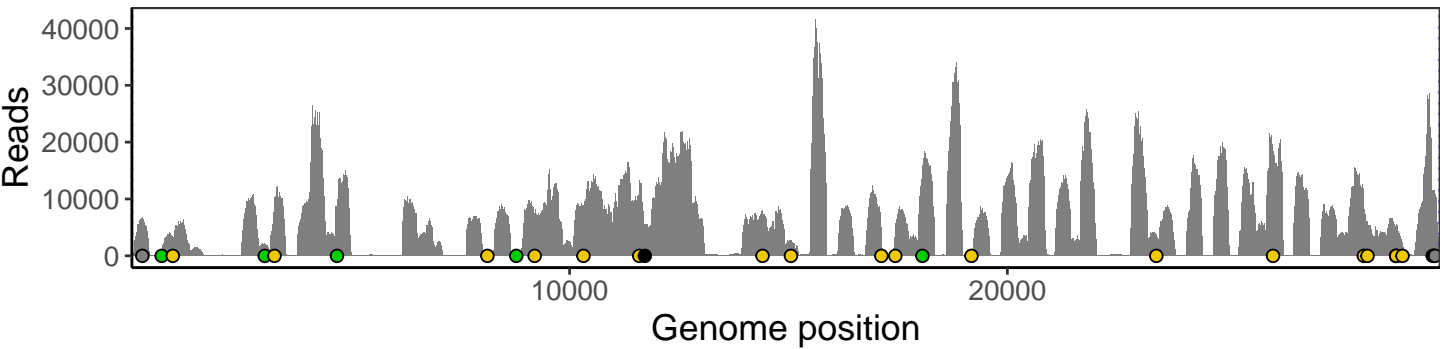
VSP0637-1

VSP0637-2

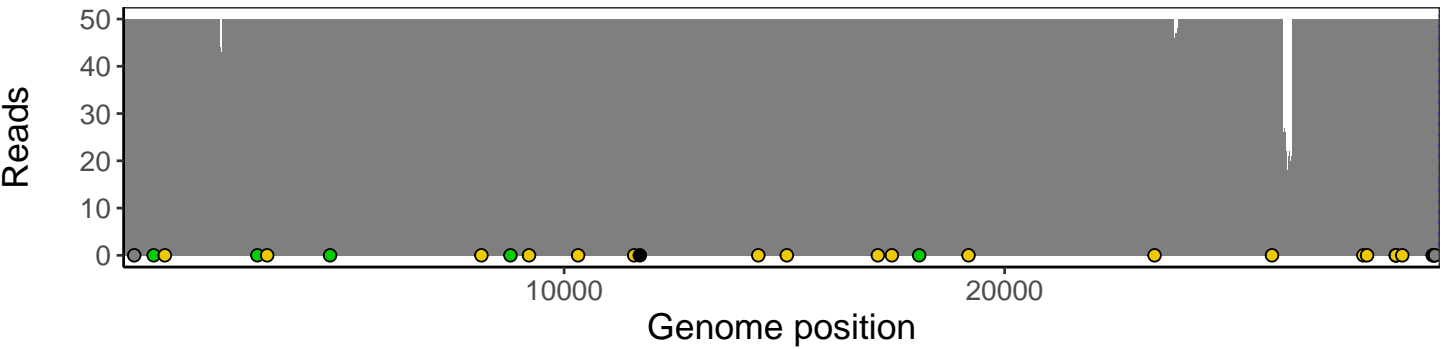
Analyses of individual experiments and composite results

VSP0637 | 2021-01-22 | VTM | H2101200627 | composite result

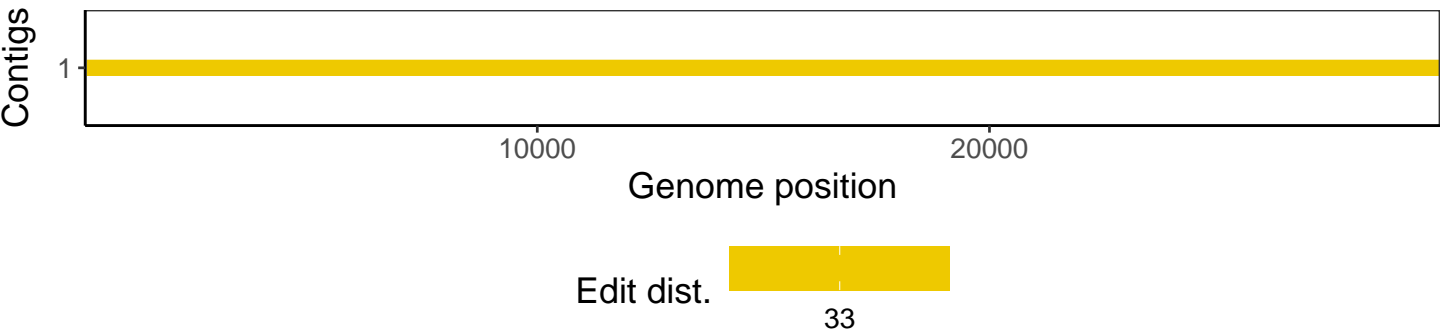
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according to variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



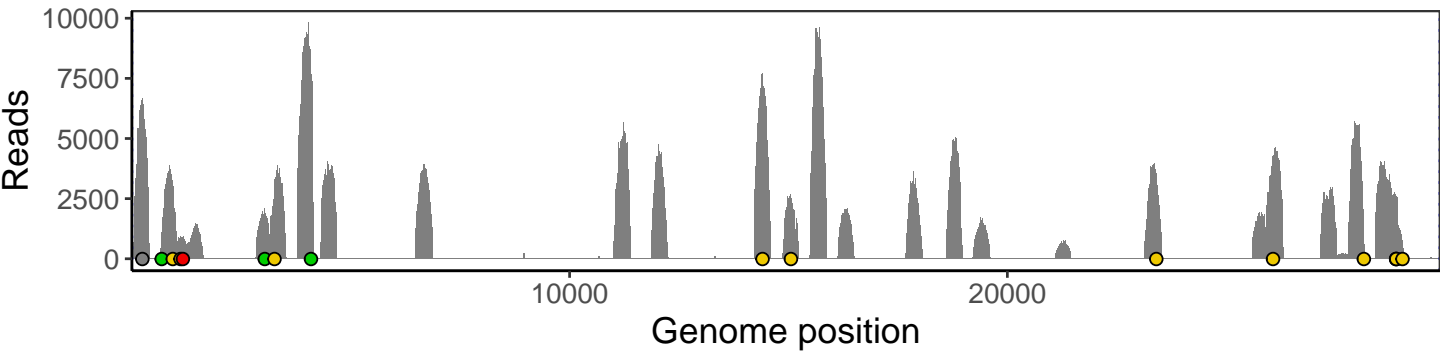
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



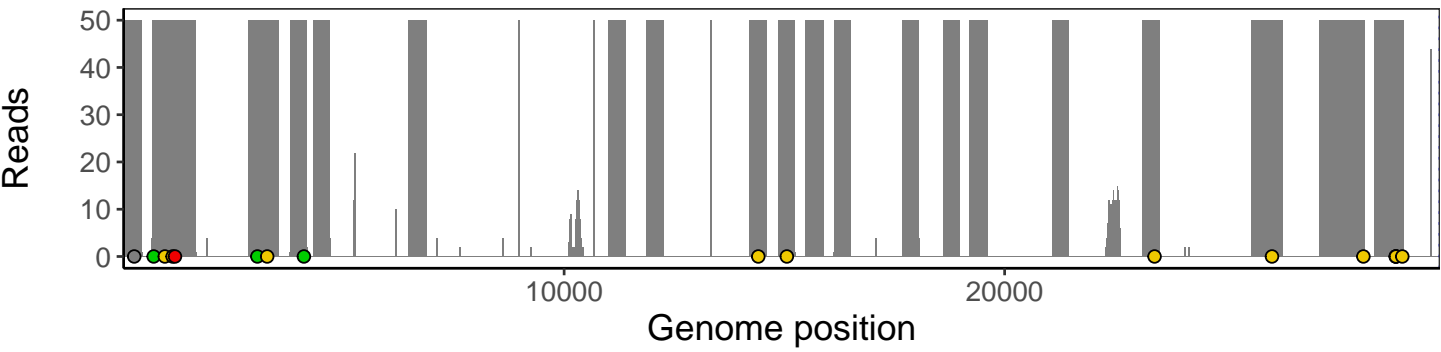
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



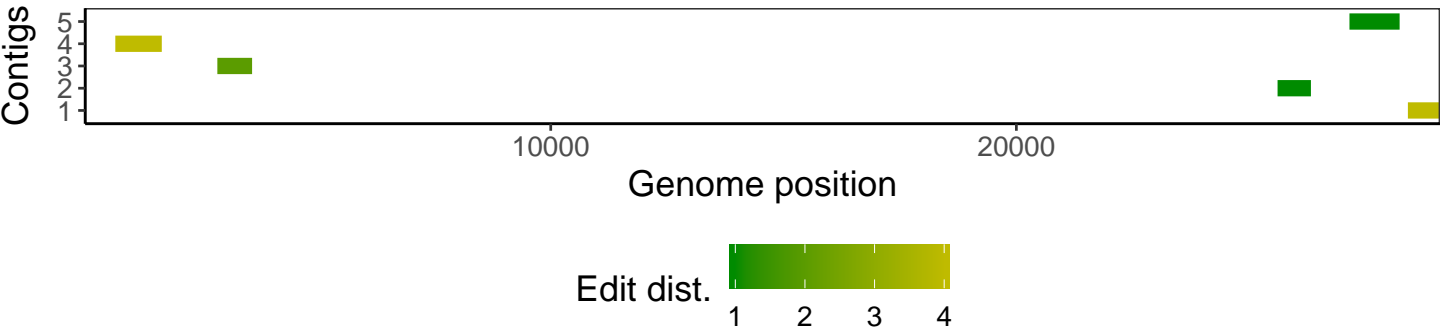
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



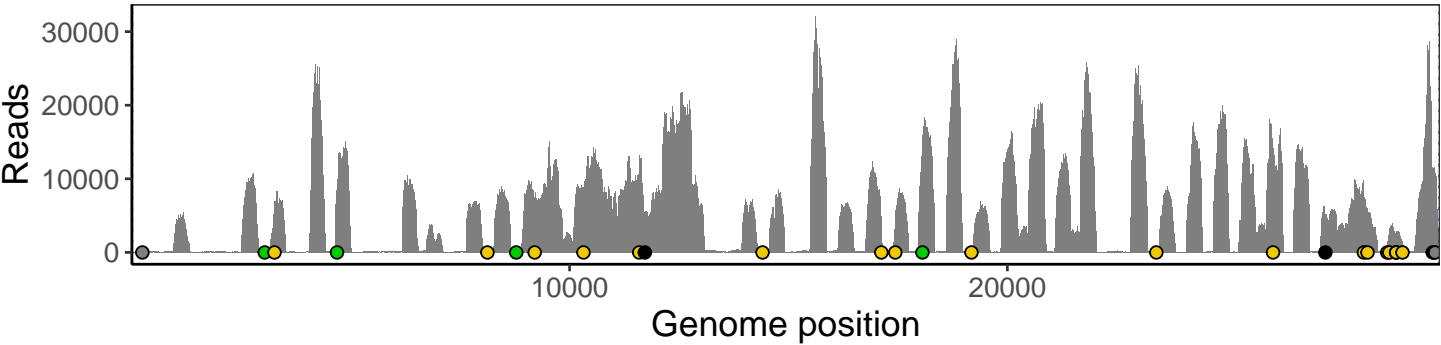
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



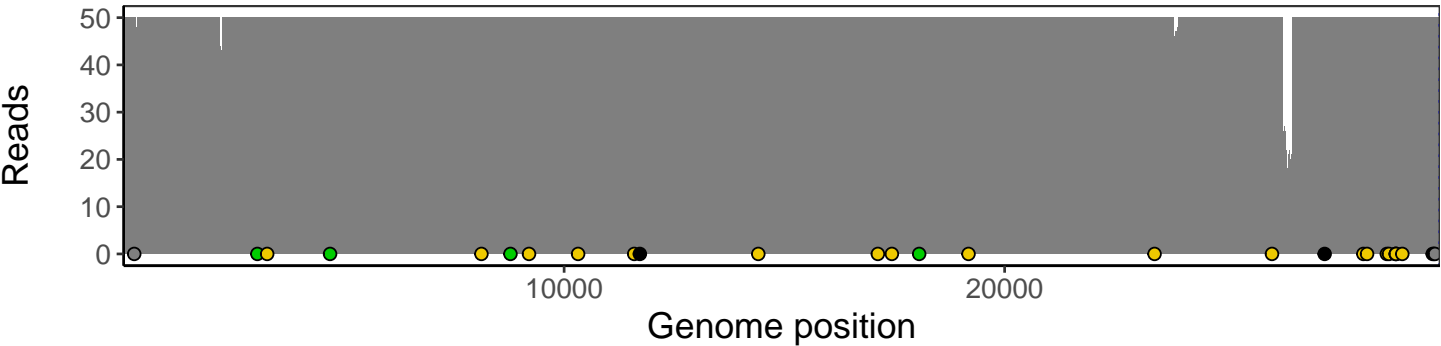
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



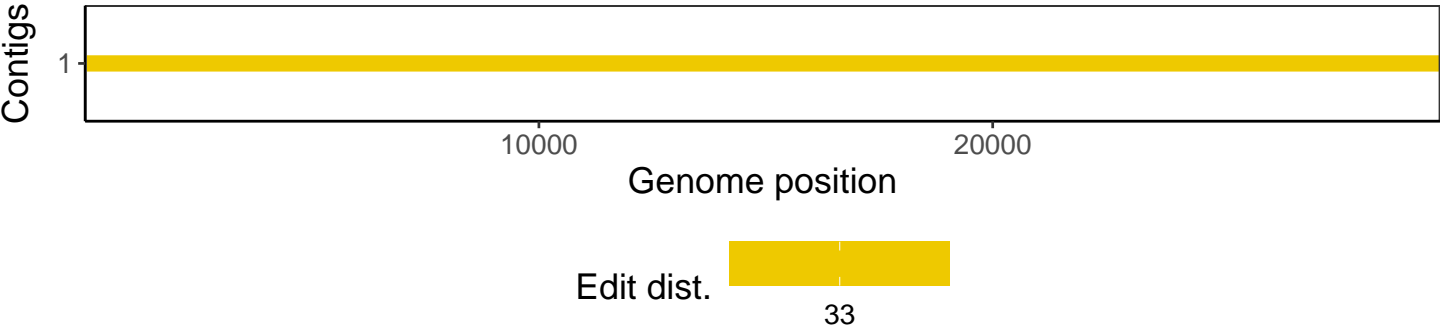
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.3.3
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