# COVID-19 subject H2103090778

2021-05-05

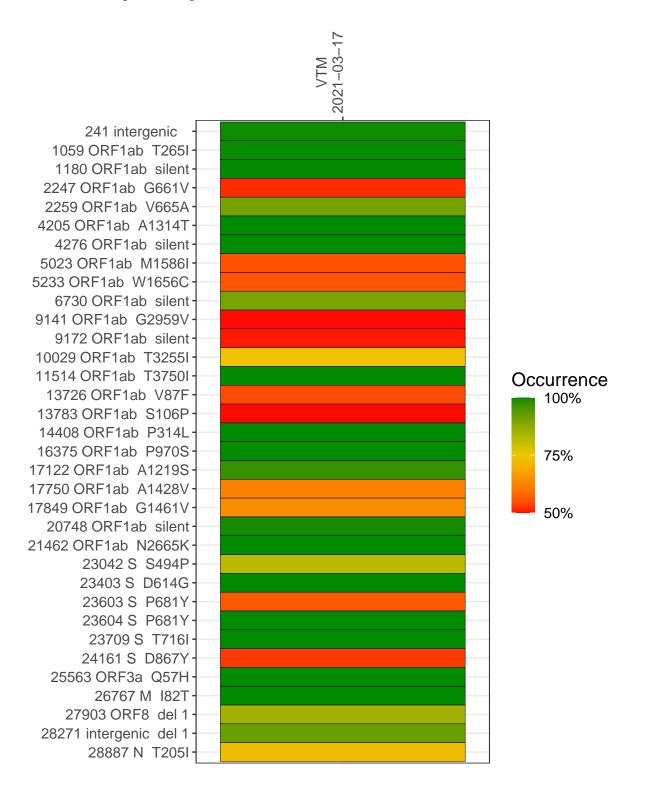
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

Table 1. Sample summary.

Experiment	Туре	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0701-1	single experiment	NA	VTM	2021-03-17	26.57	B.1.575	97.7%	97.6%

#### Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/Wuhan-Hu-1) 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.



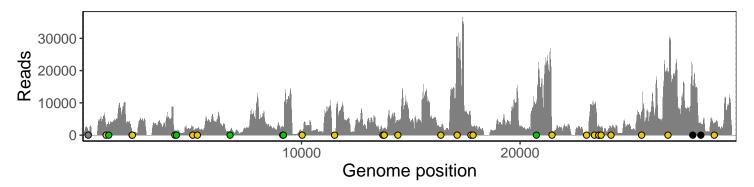
#### VTM 2021-03-17

	2021-03-17
241 intergenic	1314
1059 ORF1ab T265I	2561
1180 ORF1ab silent	3789
2247 ORF1ab G661V	1005
2259 ORF1ab V665A	664
4205 ORF1ab A1314T	1179
4276 ORF1ab silent	1439
5023 ORF1ab M1586I	2673
5233 ORF1ab W1656C	1644
6730 ORF1ab silent	1337
9141 ORF1ab G2959V	2184
9172 ORF1ab silent	2258
10029 ORF1ab T3255I	1008
11514 ORF1ab T3750I	731
13726 ORF1ab V87F	1929
13783 ORF1ab S106P	1957
14408 ORF1ab P314L	4077
16375 ORF1ab P970S	2303
17122 ORF1ab A1219S	29834
17750 ORF1ab A1428V	4149
17849 ORF1ab G1461V	2547
20748 ORF1ab silent	9820
21462 ORF1ab N2665K	1067
23042 S S494P	118
23403 S D614G	9265
23603 S P681Y	2812
23604 S P681Y	2805
23709 S T716I	2454
24161 S D867Y	2136
25563 ORF3a Q57H	<b>7227</b>
26767 M 182T	18079
27903 ORF8 del 1	13742
28271 intergenic del 1	3453
28887 N T205I	3684
	<u> </u>
	.020
	VSP0701-1
	>

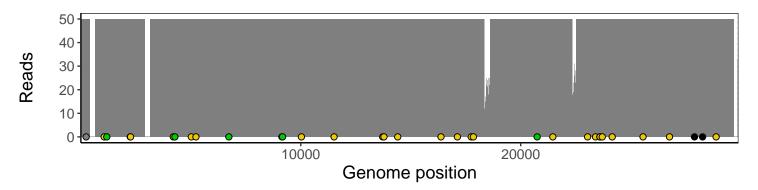
### Analyses of individual experiments and composite results

#### VSP0701-1 | 2021-03-17 | VTM | H2103090778 | genomes | single experiment

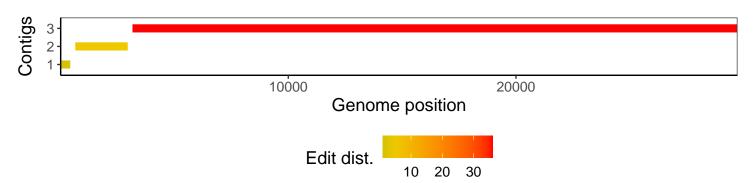
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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 htslib 1.10
bcftools	1.10.2-34-g1a12af0-dirty Using htslib 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
${\it Genomic Alignments}$	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
$\operatorname{GenomeInfoDb}$	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