COVID-19 subject UPHS-0609

2021-06-01

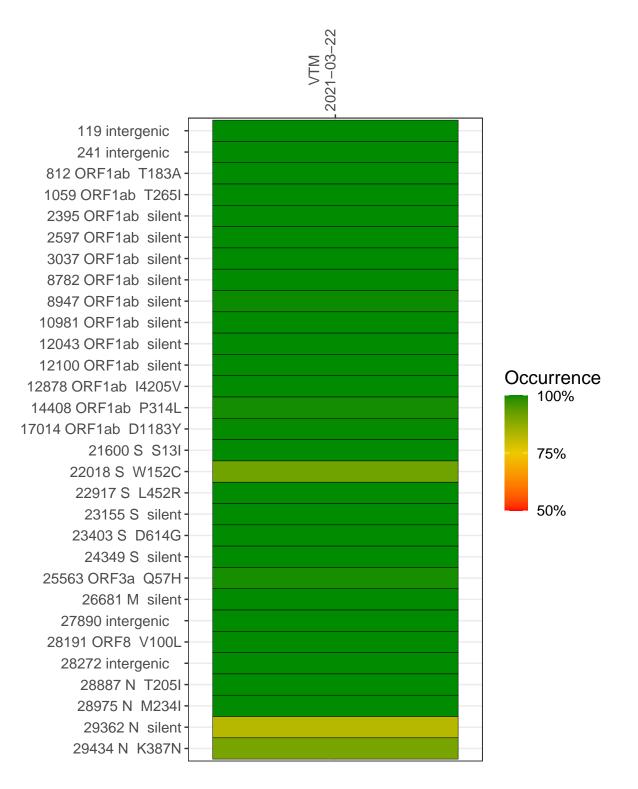
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1794-1	single experiment	NA	VTM	2021-03-22	29.83	B.1.429	99.8%	99.7%

Variants shared across samples

The heat map below shows how variants (reference genome /home/common/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 2021-03-22

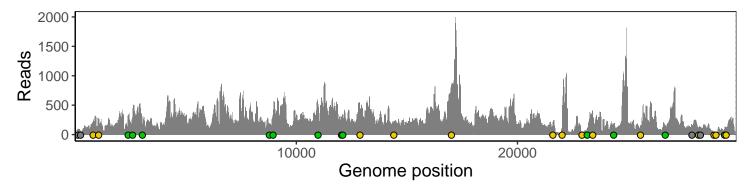
	2021 00 22
119 intergenic	93
241 intergenic	56
812 ORF1ab T183A	229
1059 ORF1ab T265I	263
2395 ORF1ab silent	339
2597 ORF1ab silent	307
3037 ORF1ab silent	207
8782 ORF1ab silent	160
8947 ORF1ab silent	363
10981 ORF1ab silent	466
12043 ORF1ab silent	295
12100 ORF1ab silent	331
12878 ORF1ab I4205V	366
14408 ORF1ab P314L	177
17014 ORF1ab D1183Y	824
21600 S S13I	211
22018 S W152C	150
22917 S L452R	46
23155 S silent	109
23403 S D614G	403
24349 S silent	161
25563 ORF3a Q57H	153
26681 M silent	135
27890 intergenic	159
28191 ORF8 V100L	122
28272 intergenic	131
28887 N T205I	31
28975 N M234I	36
29362 N silent	85
29434 N K387N	126
	1-46
	Ġ.



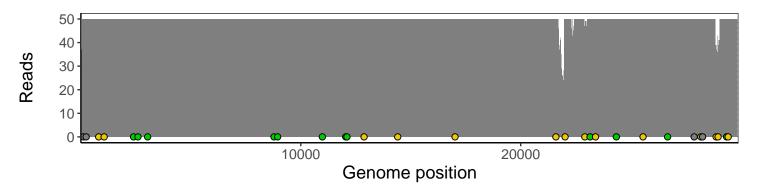
Analyses of individual experiments and composite results

VSP1794-1 | 2021-03-22 | VTM | UPHS-0609 | genomes | single experiment

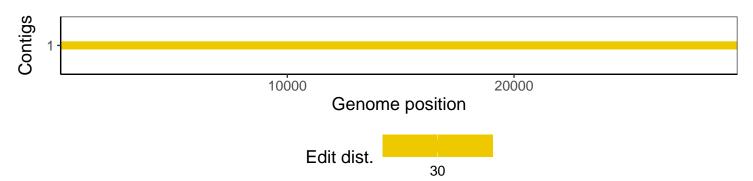
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	3.3.3
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