# COVID-19 subject UPHS-1543

2021-06-23

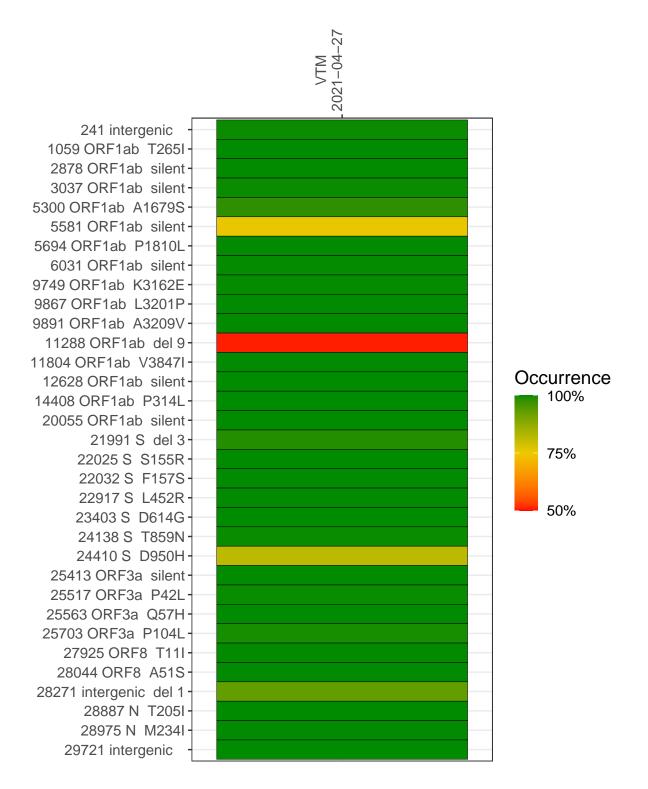
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP2840-1	single experiment	NA	VTM	2021-04-27	21.77	B.1.526	99.6%	98.8%

#### 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 2021-04-27

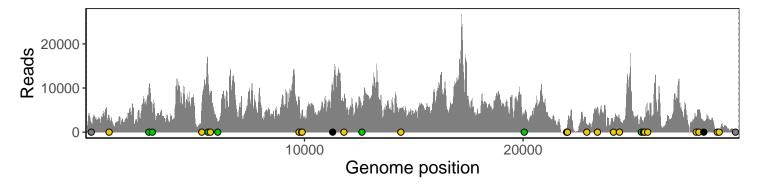
	2021–04–27
241 intergenic	1923
1059 ORF1ab T265I	3492
2878 ORF1ab silent	7460
3037 ORF1ab silent	4615
5300 ORF1ab A1679S	5554
5581 ORF1ab silent	12948
5694 ORF1ab P1810L	8472
6031 ORF1ab silent	2313
9749 ORF1ab K3162E	8074
9867 ORF1ab L3201P	3633
9891 ORF1ab A3209V	4990
11288 ORF1ab del 9	5115
11804 ORF1ab V3847I	6643
12628 ORF1ab silent	5472
14408 ORF1ab P314L	4761
20055 ORF1ab silent	4348
21991 S del 3	704
22025 S S155R	1504
22032 S F157S	1725
22917 S L452R	1715
23403 S D614G	5638
24138 S T859N	2375
24410 S D950H	2974
25413 ORF3a silent	3390
25517 ORF3a P42L	1991
25563 ORF3a Q57H	3257
25703 ORF3a P104L	4444
27925 ORF8 T11I	3312
28044 ORF8 A51S	4728
28271 intergenic del 1	2452
28887 N T205I	501
28975 N M234I	530
29721 intergenic	81
	<u> </u>
	840
	VSP2840-1
	<i>S</i>



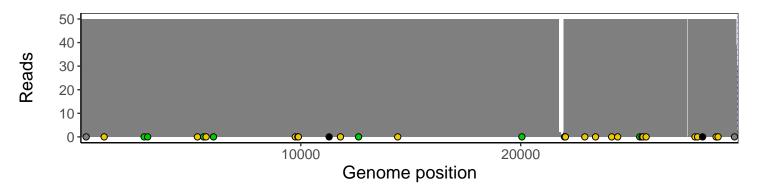
## Analyses of individual experiments and composite results

## VSP2840-1 | 2021-04-27 | VTM | UPHS-1543 | genomes | single experiment

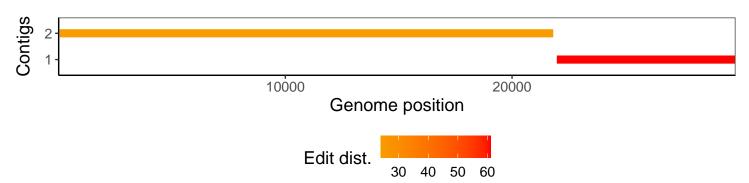
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	3.1.3				
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	3.3.3				
tidyverse	1.2.1				
ShortRead	1.34.2				
GenomicAlignments	1.12.2				
${\bf Summarized Experiment}$	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				