# COVID-19 subject UPHS-0512

2021-06-23

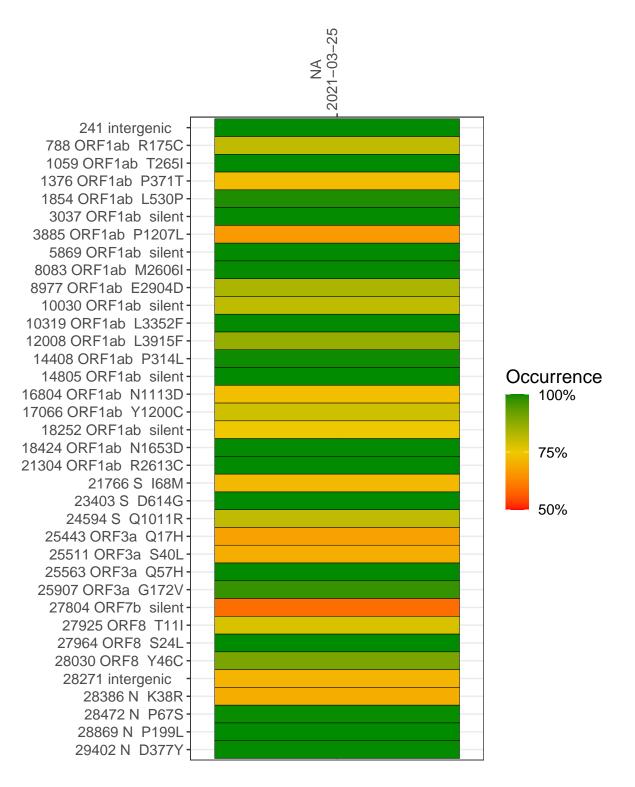
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1638-1	single experiment	NA	NA	2021-03-25	29.82	B.1.2	99.9%	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.



#### NA 2021-03-25

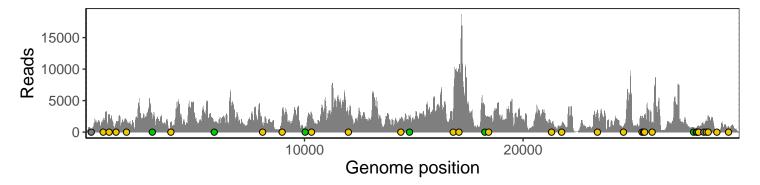
	2021 00 20
241 intergenic	<b>3</b> 63
788 ORF1ab R175C	1816
1059 ORF1ab T265I	2686
1376 ORF1ab P371T	1490
1854 ORF1ab L530P	1843
3037 ORF1ab silent	1976
3885 ORF1ab P1207L	1596
5869 ORF1ab silent	1549
8083 ORF1ab M2606I	1322
8977 ORF1ab E2904D	3354
10030 ORF1ab silent	872
10319 ORF1ab L3352F	3011
12008 ORF1ab L3915F	3124
14408 ORF1ab P314L	1812
14805 ORF1ab silent	2494
16804 ORF1ab N1113D	2745
17066 ORF1ab Y1200C	9947
18252 ORF1ab silent	4240
18424 ORF1ab N1653D	1915
21304 ORF1ab R2613C	1230
21766 S I68M	1269
23403 S D614G	3440
24594 S Q1011R	918
25443 ORF3a Q17H	2221
25511 ORF3a S40L	2231
25563 ORF3a Q57H	2733
25907 ORF3a G172V	1437
27804 ORF7b silent	759
27925 ORF8 T11I	694
27964 ORF8 S24L	813
28030 ORF8 Y46C	1038
28271 intergenic	1174
28386 N K38R	1967
28472 N P67S	2658
28869 N P199L	222
29402 N D377Y	921
	VSP1638–1
	<u></u>
	>



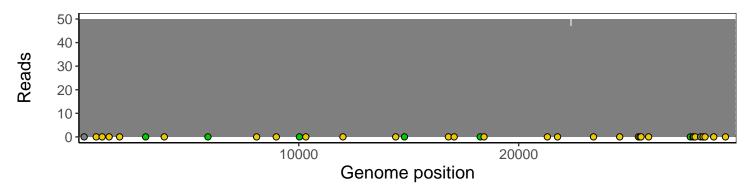
### Analyses of individual experiments and composite results

#### VSP1638-1 | 2021-03-25 | NA | UPHS-0512 | genomes | single experiment

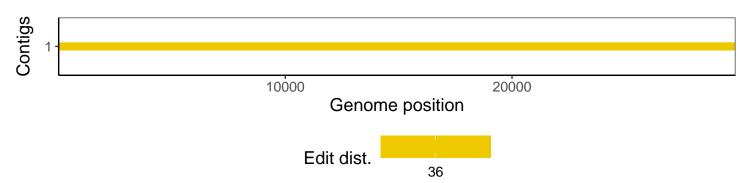
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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				