

# COVID-19 subject UPHS-0245

*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.

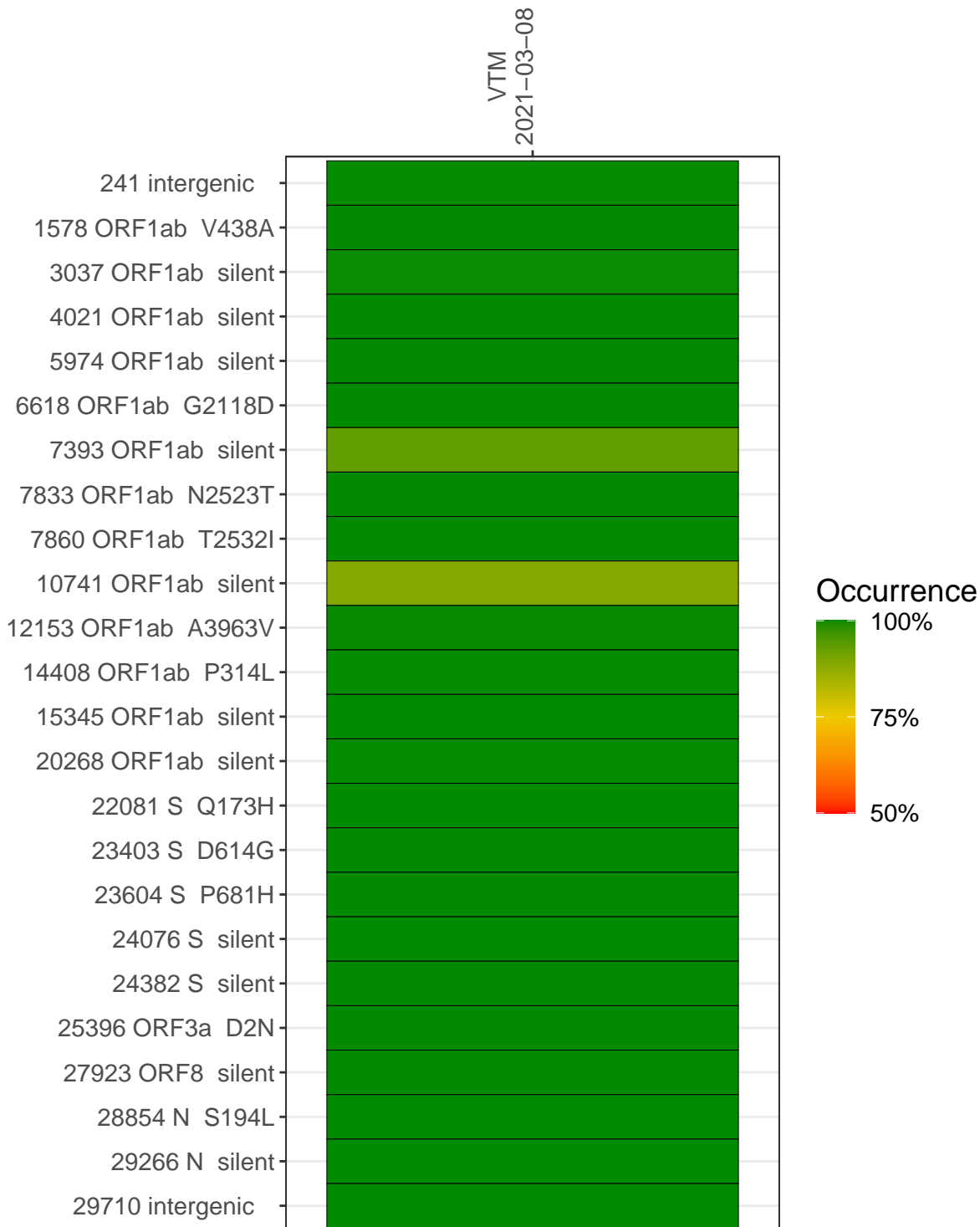
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with  $> 90\%$  sequence coverage.

Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage ( $\geq 5$ reads)
VSP1290-1	single experiment	NA	VTM	2021-03-08	29.86	B.1.243	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  
2021-03-08

241 intergenic	2261
1578 ORF1ab V438A	3792
3037 ORF1ab silent	6028
4021 ORF1ab silent	5116
5974 ORF1ab silent	4988
6618 ORF1ab G2118D	15673
7393 ORF1ab silent	14796
7833 ORF1ab N2523T	10326
7860 ORF1ab T2532I	9123
10741 ORF1ab silent	8047
12153 ORF1ab A3963V	10607
14408 ORF1ab P314L	8625
15345 ORF1ab silent	15266
20268 ORF1ab silent	3211
22081 S Q173H	6578
23403 S D614G	9144
23604 S P681H	8901
24076 S silent	4076
24382 S silent	6871
25396 ORF3a D2N	4828
27923 ORF8 silent	7880
28854 N S194L	563
29266 N silent	3322
29710 intergenic	7079

Base change

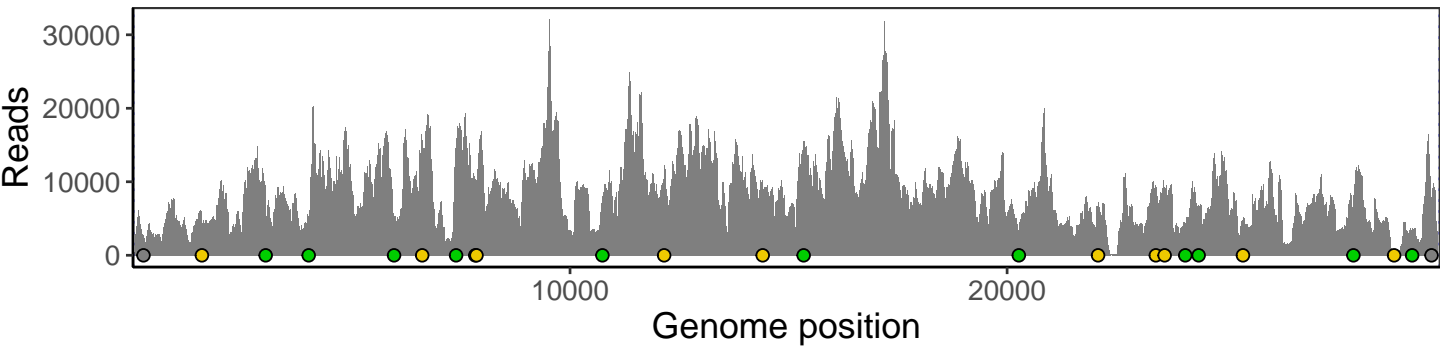
- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

VSP1290-1

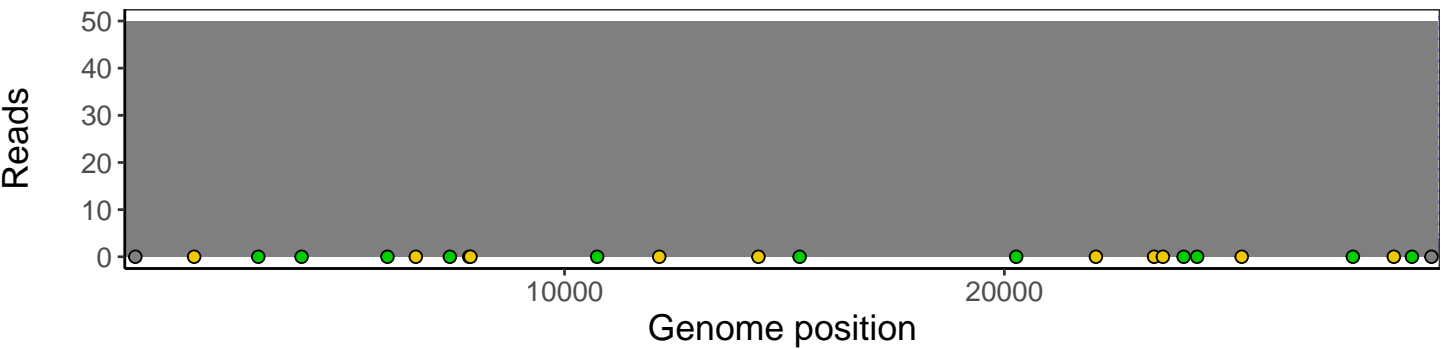
# Analyses of individual experiments and composite results

VSP1290-1 | 2021-03-08 | VTM | UPHS-0245 | genomes | single experiment

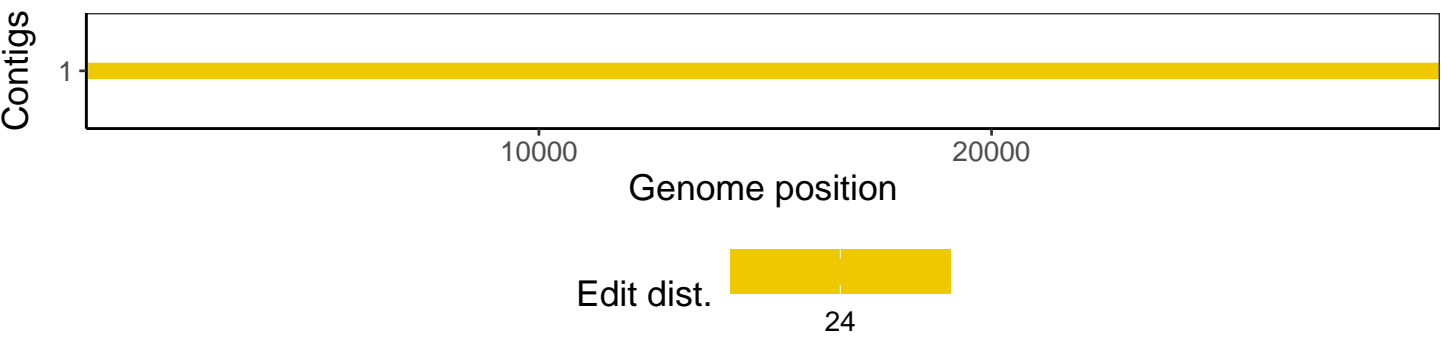
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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 htlib 1.10
bcftools	1.10.2-34-g1a12af0-dirty Using htlib 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
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