# COVID-19 subject UPHS-1376

2021-06-23

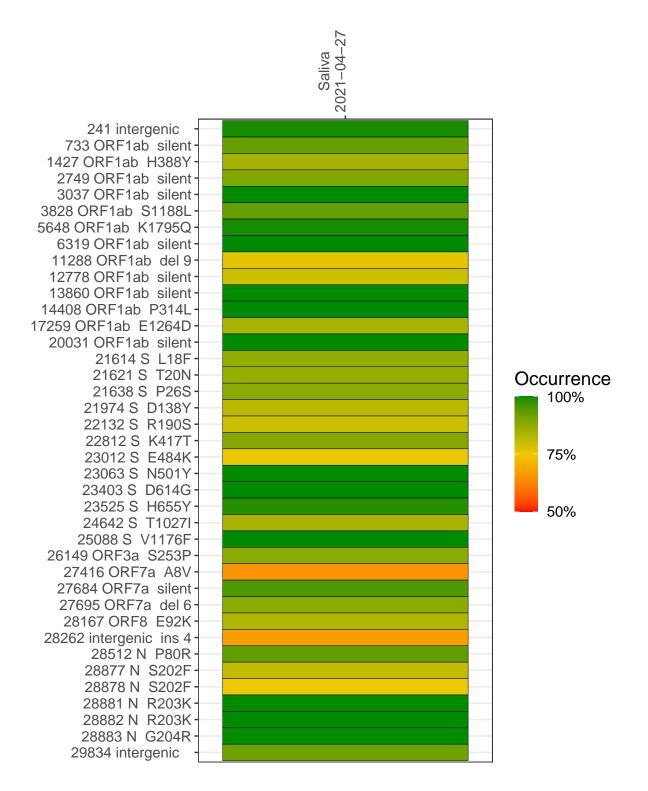
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP2631-1	single experiment	NA	Saliva	2021-04-27	23.11	P.1	99.3%	99.2%

#### 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.



### Saliva 2021-04-27

Base change Expected

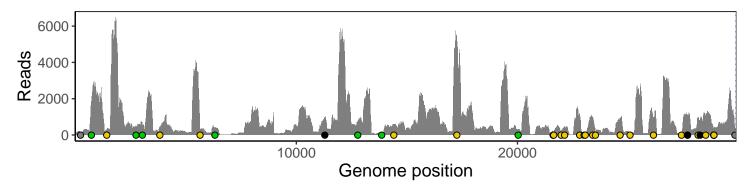
Ins/Del No data

	2021-04-27
241 intergenic	152
733 ORF1ab silent	1830
1427 ORF1ab H388Y	294
2749 ORF1ab silent	431
3037 ORF1ab silent	541
3828 ORF1ab S1188L	445
5648 ORF1ab K1795Q	294
6319 ORF1ab silent	357
11288 ORF1ab del 9	621
12778 ORF1ab silent	738
13860 ORF1ab silent	124
14408 ORF1ab P314L	530
17259 ORF1ab E1264D	5030
20031 ORF1ab silent	295
21614 S L18F	178
21621 S T20N	156
21638 S P26S	156
21974 S D138Y	204
22132 S R190S	458
22812 S K417T	1062
23012 S E484K	89
23063 S N501Y	101
23403 S D614G	941
23525 S H655Y	139
24642 S T1027I	876
25088 S V1176F	91
26149 ORF3a S253P	890
27416 ORF7a A8V	109
27684 ORF7a silent	1074
27695 ORF7a del 6	956
28167 ORF8 E92K	850
28262 intergenic ins 4	652
28512 N P80R	983
28877 N S202F	239
28878 N S202F	238
28881 N R203K	225
28882 N R203K	225
28883 N G204R	226
29834 intergenic	379
•	
	2631–1
	56

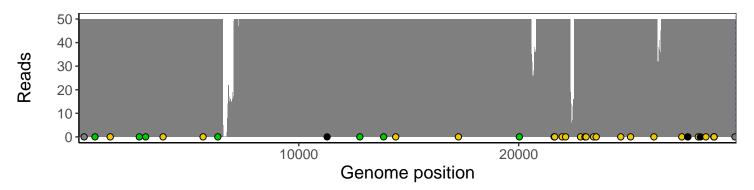
## Analyses of individual experiments and composite results

## VSP2631-1 | 2021-04-27 | Saliva | UPHS-1376 | genomes | single experiment

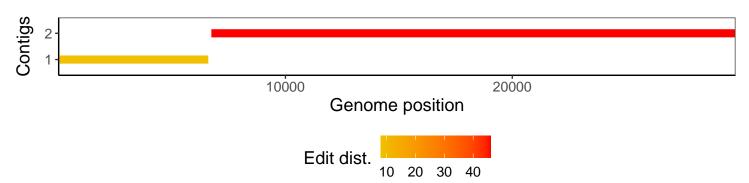
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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