# COVID-19 subject UPHS-0705

2021-05-21

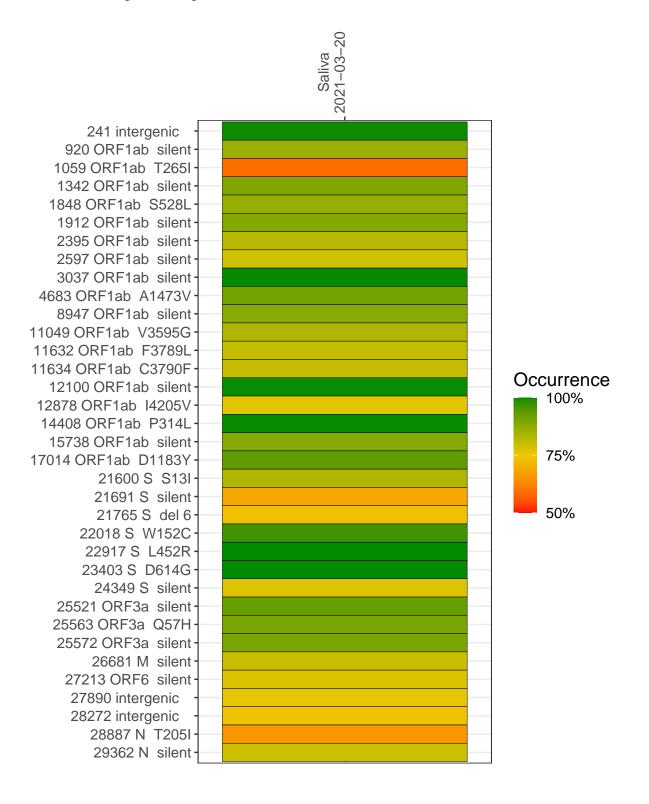
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1923-2	single experiment	NA	Saliva	2021-03-20	29.67	B.1.429	99.3%	99.3%

#### Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/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 2021–03–20

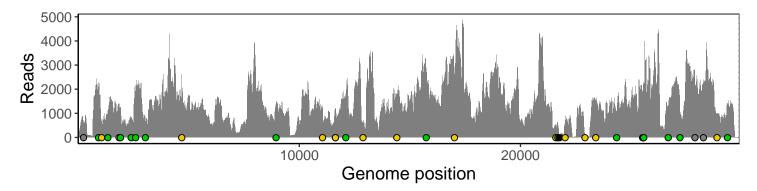
	2021 00 20
241 intergenic	593
920 ORF1ab silent	2131
1059 ORF1ab T265I	562
1342 ORF1ab silent	1094
1848 ORF1ab S528L	1180
1912 ORF1ab silent	1137
2395 ORF1ab silent	805
2597 ORF1ab silent	1564
3037 ORF1ab silent	672
4683 ORF1ab A1473V	1548
8947 ORF1ab silent	1376
11049 ORF1ab V3595G	1170
11632 ORF1ab F3789L	1149
11634 ORF1ab C3790F	1149
12100 ORF1ab silent	1655
12878 ORF1ab I4205V	641
14408 ORF1ab P314L	1628
15738 ORF1ab silent	2701
17014 ORF1ab D1183Y	2526
21600 S S13I	518
21691 S silent	532
21765 S del 6	108
22018 S W152C	627
22917 S L452R	27
23403 S D614G	1514
24349 S silent	953
25521 ORF3a silent	1850
25563 ORF3a Q57H	2212
25572 ORF3a silent	2091
26681 M silent	1106
27213 ORF6 silent	1039
27890 intergenic	2112
28272 intergenic	2091
28887 N T205I	963
29362 N silent	1355
	3-2
	VSP1923-2
	74
	>



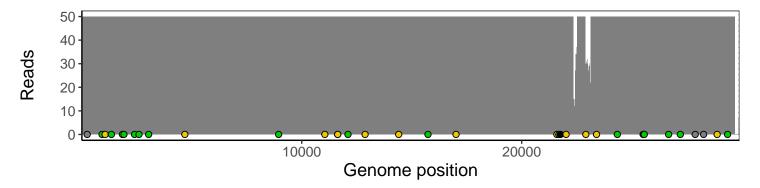
## Analyses of individual experiments and composite results

## VSP1923-2 | 2021-03-20 | Saliva | UPHS-0705 | genomes | single experiment

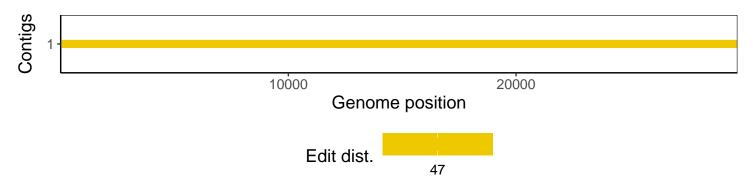
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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