

# COVID-19 subject UPHS-1338

*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.

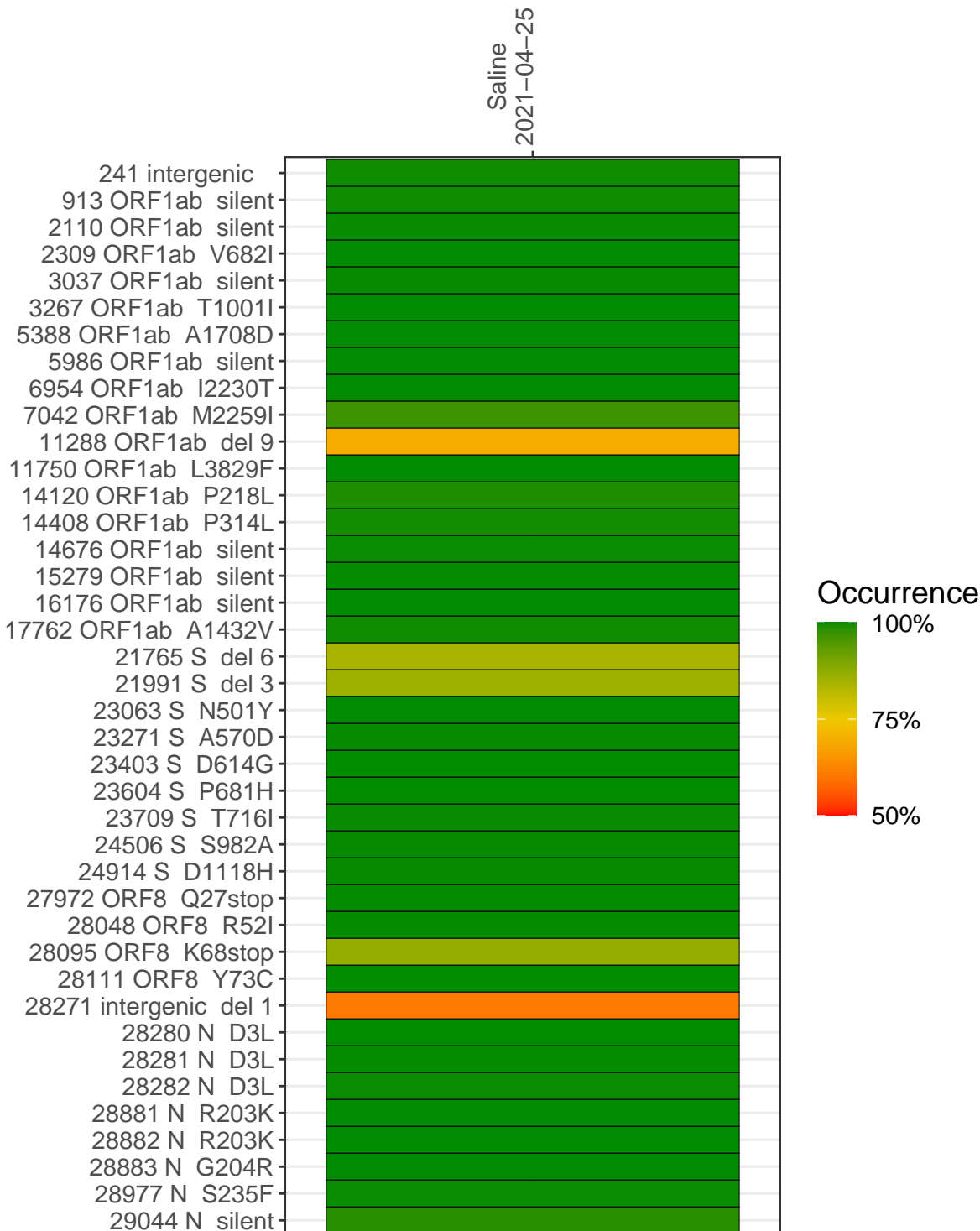
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with  $> 90\%$  sequence coverage.

Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage ( $\geq 5$ reads)
VSP2593-1	single experiment	NA	Saline	2021-04-25	29.81	B.1.1.7	99.8%	99.8%

## 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.



	Saline 2021-04-25	
241 intergenic	2607	
913 ORF1ab silent	6724	
2110 ORF1ab silent	4472	
2309 ORF1ab V682I	2324	
3037 ORF1ab silent	4475	
3267 ORF1ab T1001I	5088	
5388 ORF1ab A1708D	5555	
5986 ORF1ab silent	2629	
6954 ORF1ab I2230T	1962	
7042 ORF1ab M2259I	2631	
11288 ORF1ab del 9	3670	
11750 ORF1ab L3829F	6164	
14120 ORF1ab P218L	7240	
14408 ORF1ab P314L	5168	
14676 ORF1ab silent	4006	
15279 ORF1ab silent	7673	
16176 ORF1ab silent	6647	
17762 ORF1ab A1432V	2145	
21765 S del 6	3128	
21991 S del 3	1469	
23063 S N501Y	1383	
23271 S A570D	5911	
23403 S D614G	6221	
23604 S P681H	6293	
23709 S T716I	5511	
24506 S S982A	3267	
24914 S D1118H	6641	
27972 ORF8 Q27stop	7986	
28048 ORF8 R52I	6836	
28095 ORF8 K68stop	6309	
28111 ORF8 Y73C	5983	
28271 intergenic del 1	4128	
28280 N D3L	2492	
28281 N D3L	2493	
28282 N D3L	2701	
28881 N R203K	441	
28882 N R203K	440	
28883 N G204R	440	
28977 N S235F	492	
29044 N silent	1491	

Base change

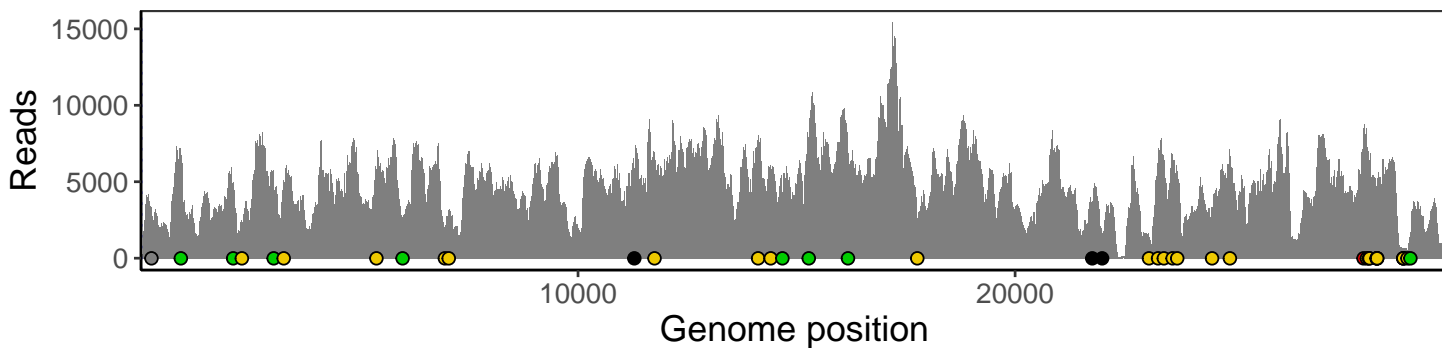
- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

VSP2593-1

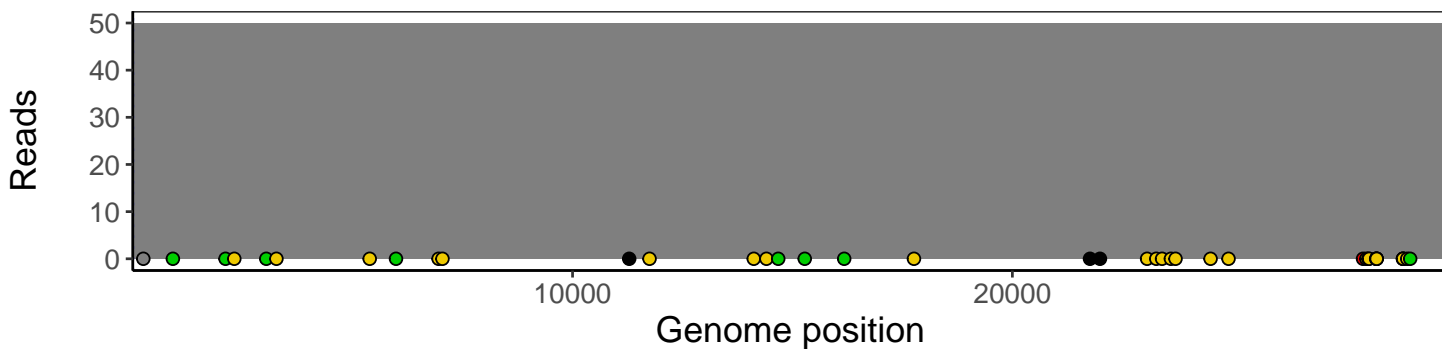
## Analyses of individual experiments and composite results

VSP2593-1 | 2021-04-25 | Saline | UPHS-1338 | genomes | single experiment

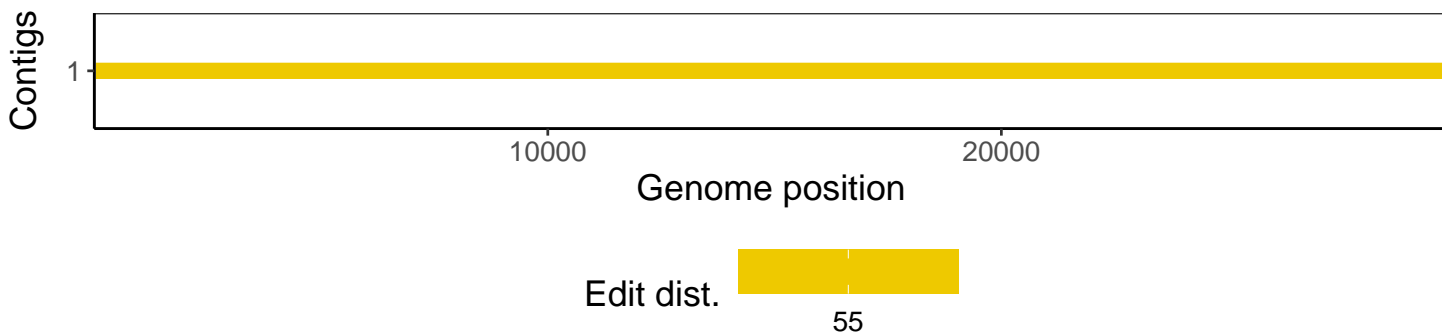
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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