

# COVID-19 subject UPHS-0074

*2021-05-05*

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.

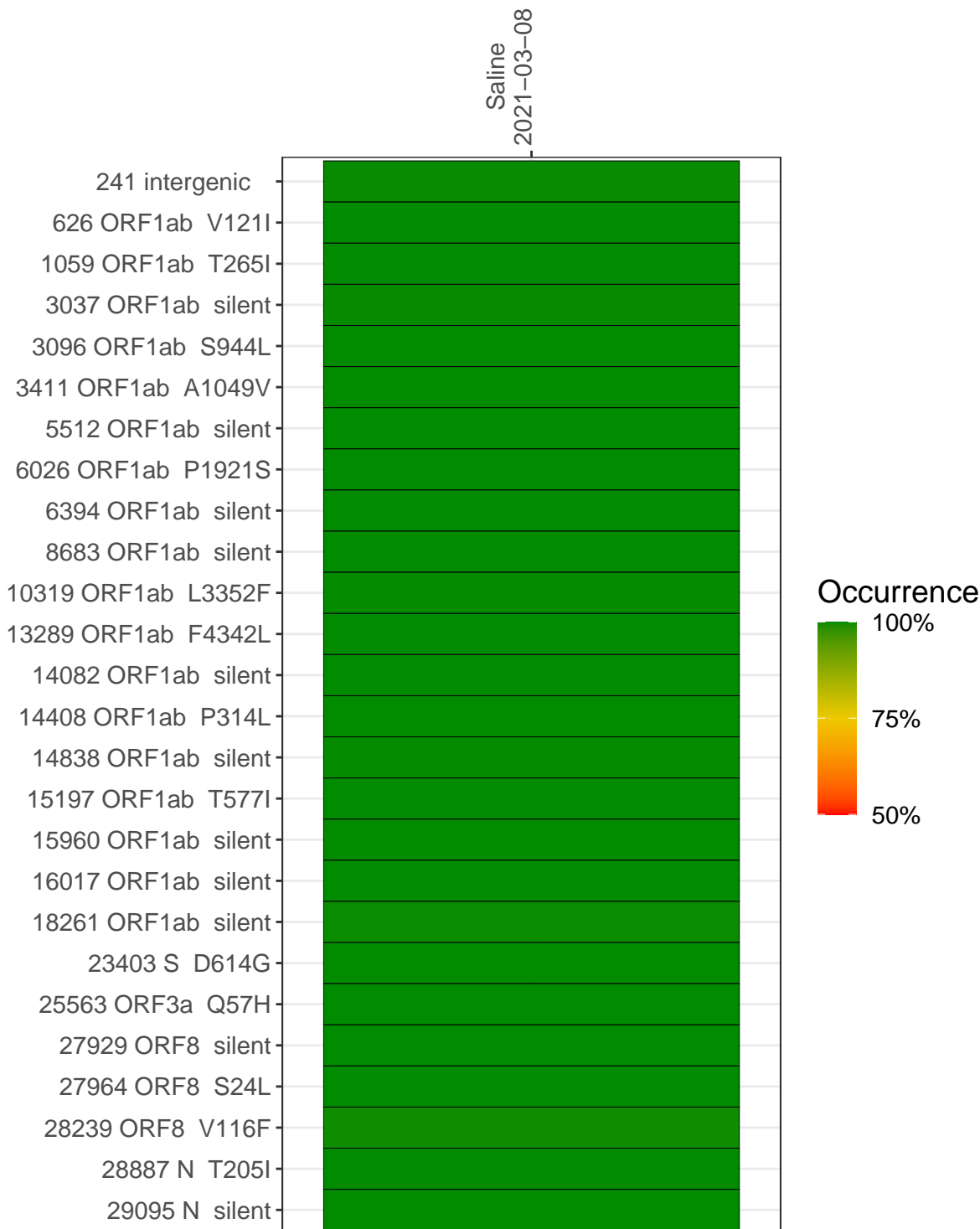
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with  $> 90\%$  sequence coverage.

Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage ( $\geq 5$ reads)
VSP1006-1	single experiment	NA	Saline	2021-03-08	29.84	B.1.595	99.8%	99.7%

## 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.



Saline  
2021-03-08

241 intergenic	2264
626 ORF1ab V121I	1044
1059 ORF1ab T265I	4395
3037 ORF1ab silent	5914
3096 ORF1ab S944L	7904
3411 ORF1ab A1049V	5001
5512 ORF1ab silent	10731
6026 ORF1ab P1921S	5190
6394 ORF1ab silent	9089
8683 ORF1ab silent	4343
10319 ORF1ab L3352F	10434
13289 ORF1ab F4342L	15026
14082 ORF1ab silent	6064
14408 ORF1ab P314L	10024
14838 ORF1ab silent	8619
15197 ORF1ab T577I	8986
15960 ORF1ab silent	16401
16017 ORF1ab silent	16284
18261 ORF1ab silent	7746
23403 S D614G	10366
25563 ORF3a Q57H	6889
27929 ORF8 silent	15374
27964 ORF8 S24L	14908
28239 ORF8 V116F	6585
28887 N T205I	1156
29095 N silent	7791

Base change

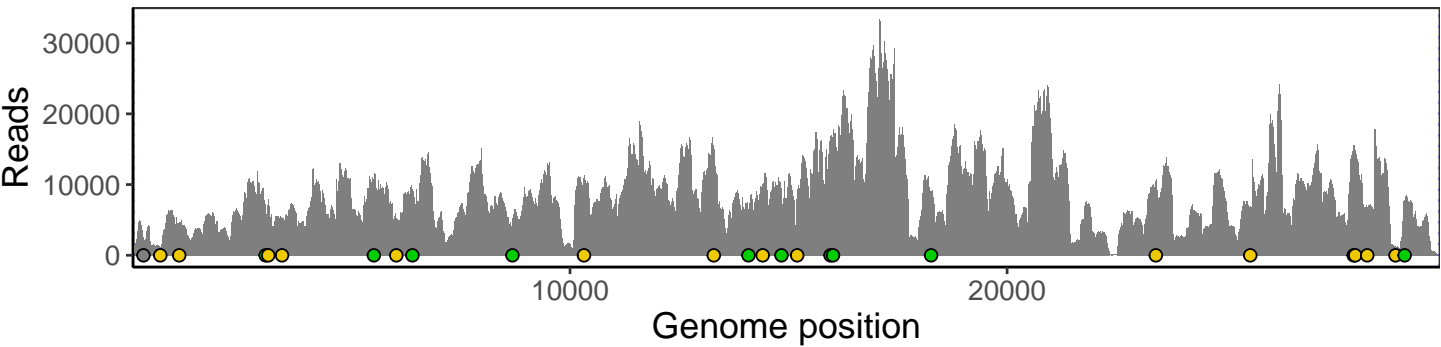
- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

VSP1006-1

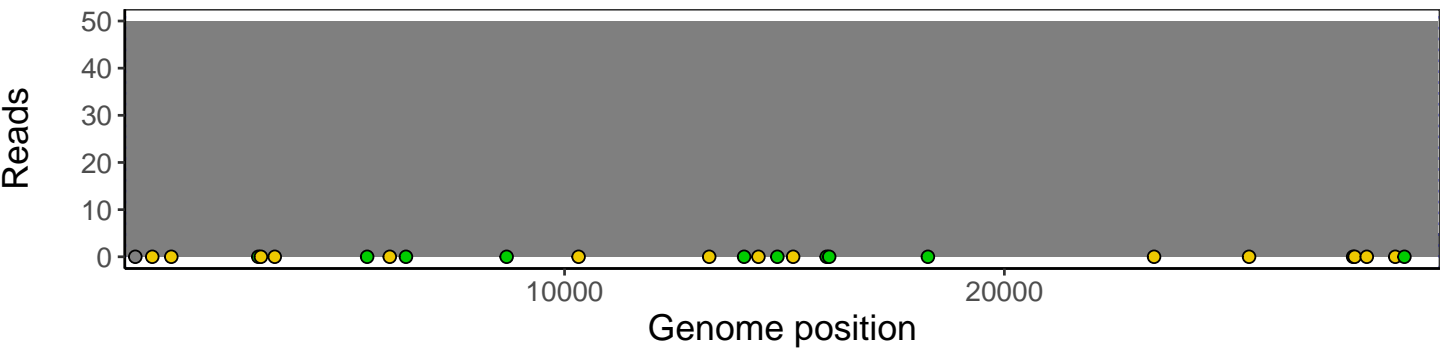
# Analyses of individual experiments and composite results

VSP1006-1 | 2021-03-08 | Saline | UPHS-0074 | genomes | single experiment

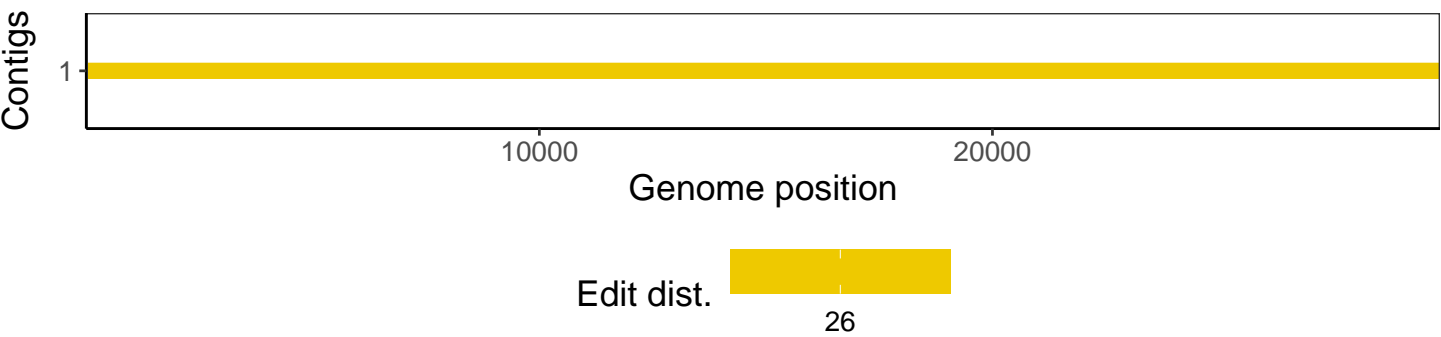
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	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	3.0.0
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