# COVID-19 subject UPHS-0789

2021-06-23

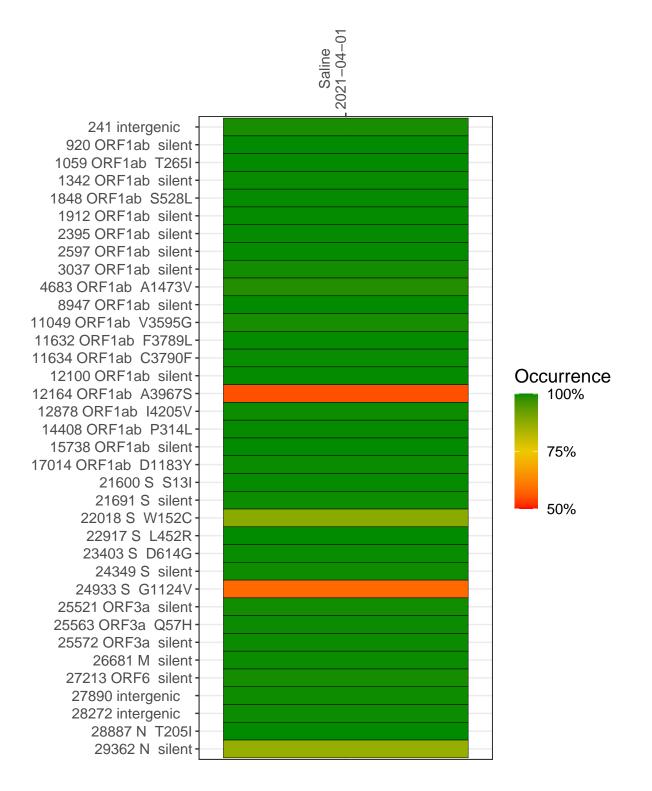
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1969-2	single experiment	NA	Saline	2021-04-01	29.84	B.1.429	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.



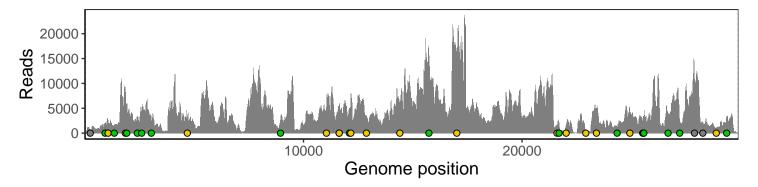
#### Saline 2021-04-01

241 intergenic	654
920 ORF1ab silent	1958
1059 ORF1ab T265I	2374
1342 ORF1ab silent	1683
1848 ORF1ab S528L	8469
1912 ORF1ab silent	5170
2395 ORF1ab silent	3304
2597 ORF1ab silent	3512
3037 ORF1ab silent	2621
4683 ORF1ab A1473V	2842
8947 ORF1ab silent	2664
11049 ORF1ab V3595G	5071
11632 ORF1ab F3789L	5703
11634 ORF1ab C3790F	5650
12100 ORF1ab silent	4865
12164 ORF1ab A3967S	7674
12878 ORF1ab I4205V	3081
14408 ORF1ab P314L	5266
15738 ORF1ab silent	15721
17014 ORF1ab D1183Y	15343
21600 S S13I	1693
21691 S silent	2085
22018 S W152C	1676
22917 S L452R	400
23403 S D614G	4906
24349 S silent	1196
24933 S G1124V	4926
25521 ORF3a silent	2353
25563 ORF3a Q57H	3332
25572 ORF3a silent	3245
26681 M silent	3379
27213 ORF6 silent	2578
27890 intergenic	10786
28272 intergenic	2142
28887 N T205I	1001
29362 N silent	3060
	VSP1969-2
	2P
	**************************************

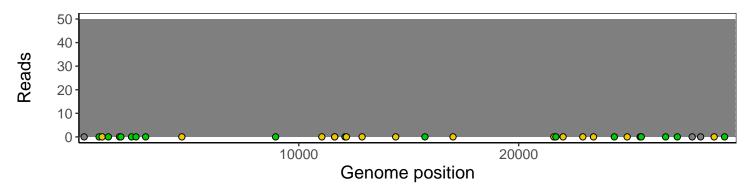
### Analyses of individual experiments and composite results

#### $VSP1969-2\mid 2021-04-01\mid Saline\mid UPHS-0789\mid genomes\mid single\ experiment$

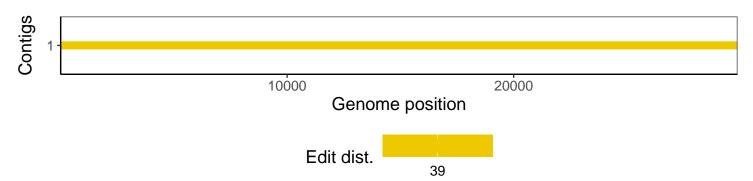
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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