# COVID-19 subject HUP Q-0154

2021-05-05

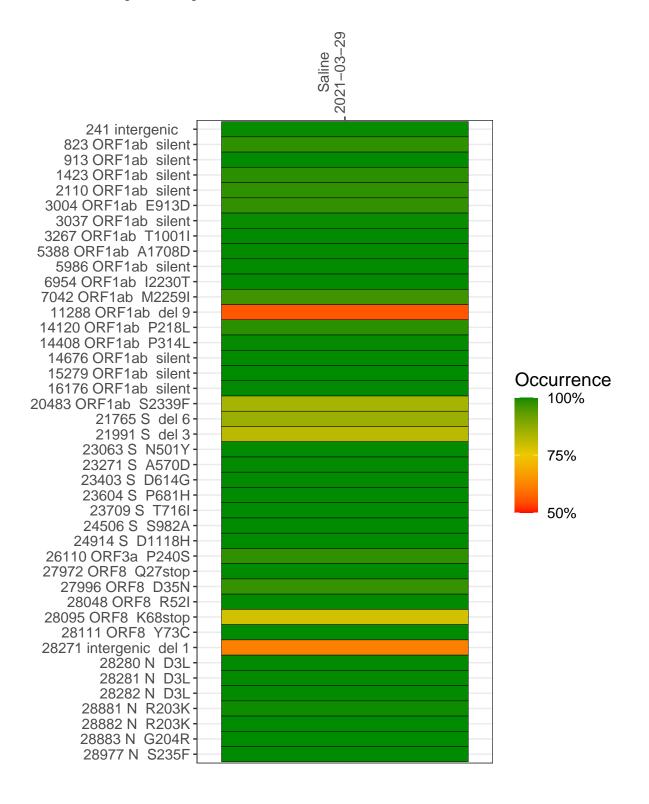
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with > 90% sequence coverage.

Table 1. Sample summary.

Experiment	Туре	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1495-1	single experiment	NA	Saline	2021-03-29	29.82	B.1.1.7	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 2021-03-29

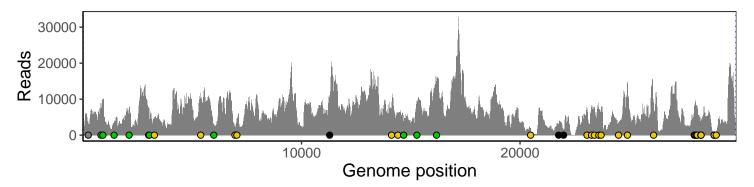
	2021-03-29
241 intergenic	3093
823 ORF1ab silent	8871
913 ORF1ab silent	9525
1423 ORF1ab silent	3278
2110 ORF1ab silent	5750
3004 ORF1ab E913D	7341
3037 ORF1ab silent	4734
3267 ORF1ab T1001I	5014
5388 ORF1ab A1708D	10008
5986 ORF1ab silent	3089
6954 ORF1ab I2230T	1756
7042 ORF1ab M2259I	3681
11288 ORF1ab del 9	5421
14120 ORF1ab P218L	7271
14408 ORF1ab P314L	5398
14676 ORF1ab silent	3088
15279 ORF1ab silent	8262
16176 ORF1ab silent	14558
20483 ORF1ab S2339F	1930
21765 S del 6	2659
21991 S del 3	1347
23063 S N501Y	7306
23271 S A570D	7453
23403 S D614G	7646
23604 S P681H	7322
23709 S T716I	6709
24506 S S982A	3461
24914 S D1118H	14650
26110 ORF3a P240S	8925
27972 ORF8 Q27stop	7725
27996 ORF8 D35N	7591
28048 ORF8 R52I	8359
28095 ORF8 K68stop	8042
28111 ORF8 Y73C	6642
28271 intergenic del 1	3579
28280 N D3L	2152
28281 N D3L	2152
28282 N D3L	2357
28881 N R203K	592
28882 N R203K	590
28883 N G204R	593
28977 N S235F	868
	7
	SP1495-1
	Ž 4
	<u>C</u>



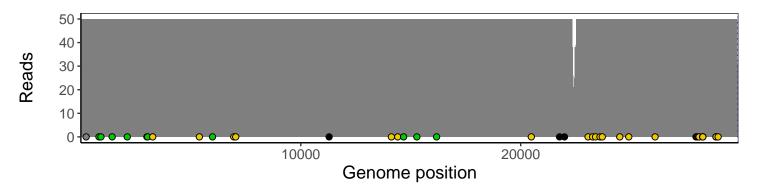
## Analyses of individual experiments and composite results

### VSP1495-1 | 2021-03-29 | Saline | HUP Q-0154 | genomes | single experiment

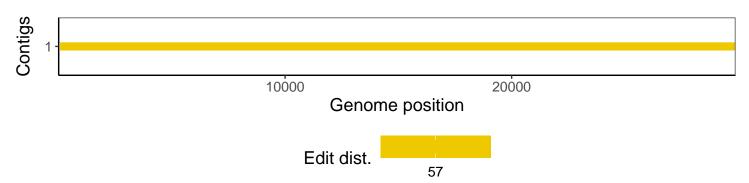
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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