

# COVID-19 subject HUP Q-0045

*2021-06-23*

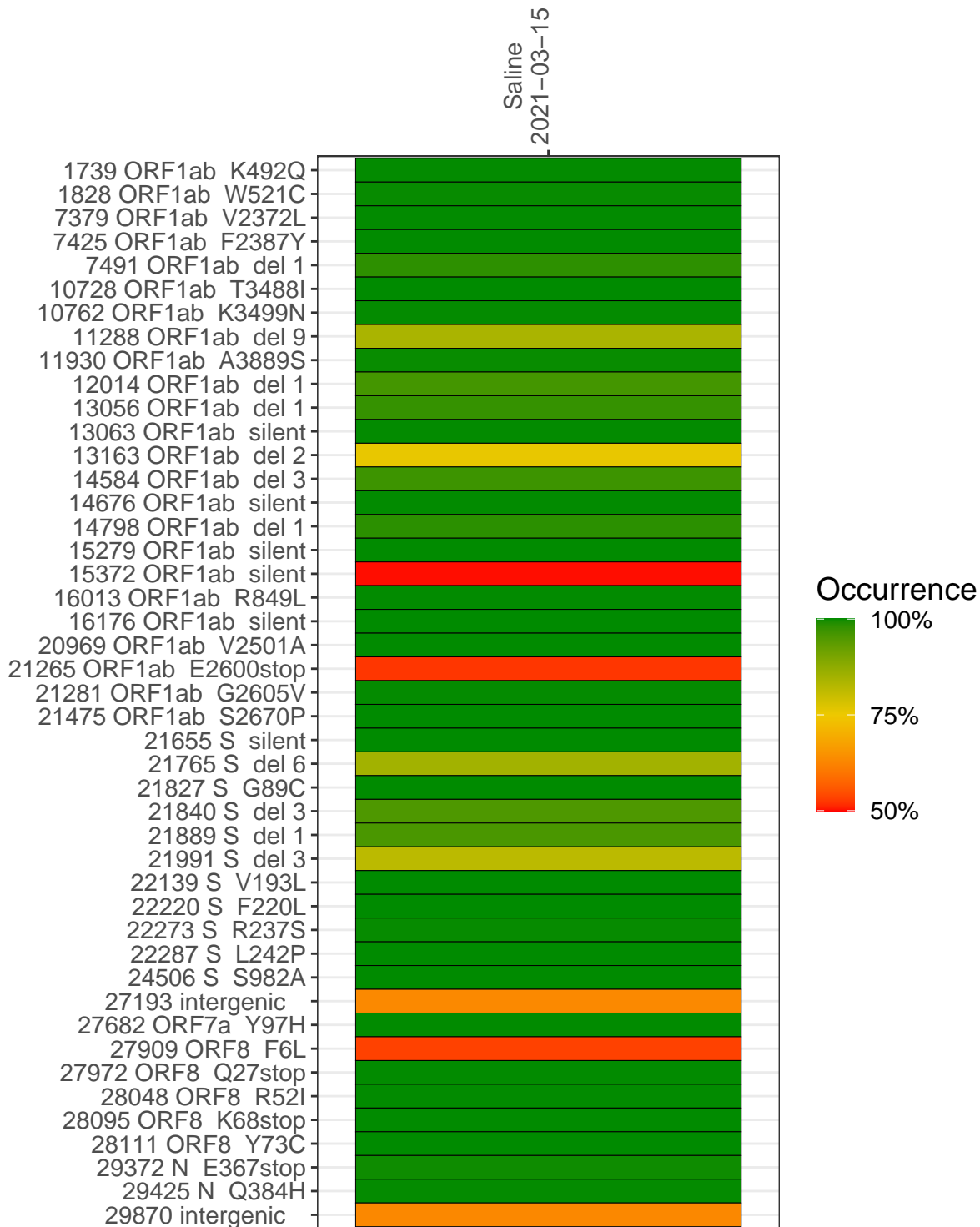
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1077-1	single experiment	NA	Saline	2021-03-15	1.53	NA	27.5%	25.6%

## 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  
2021-03-15

1739 ORF1ab K492Q	7339
1828 ORF1ab W521C	7876
7379 ORF1ab V2372L	5463
7425 ORF1ab F2387Y	7122
7491 ORF1ab del 1	9098
10728 ORF1ab T3488I	2701
10762 ORF1ab K3499N	3526
11288 ORF1ab del 9	5559
11930 ORF1ab A3889S	8595
12014 ORF1ab del 1	10299
13056 ORF1ab del 1	8181
13063 ORF1ab silent	8918
13163 ORF1ab del 2	9097
14584 ORF1ab del 3	2679
14676 ORF1ab silent	4110
14798 ORF1ab del 1	6907
15279 ORF1ab silent	16830
15372 ORF1ab silent	23048
16013 ORF1ab R849L	9123
16176 ORF1ab silent	4035
20969 ORF1ab V2501A	5938
21265 ORF1ab E2600stop	15157
21281 ORF1ab G2605V	15696
21475 ORF1ab S2670P	218
21655 S silent	171
21765 S del 6	7995
21827 S G89C	9564
21840 S del 3	8822
21889 S del 1	8460
21991 S del 3	3483
22139 S V193L	1408
22220 S F220L	1672
22273 S R237S	747
22287 S L242P	688
24506 S S982A	5714
27193 intergenic	903
27682 ORF7a Y97H	4461
27909 ORF8 F6L	31068
27972 ORF8 Q27stop	36415
28048 ORF8 R52I	30377
28095 ORF8 K68stop	25459
28111 ORF8 Y73C	20839
29372 N E367stop	2010
29425 N Q384H	3388
29870 intergenic	46

Base change

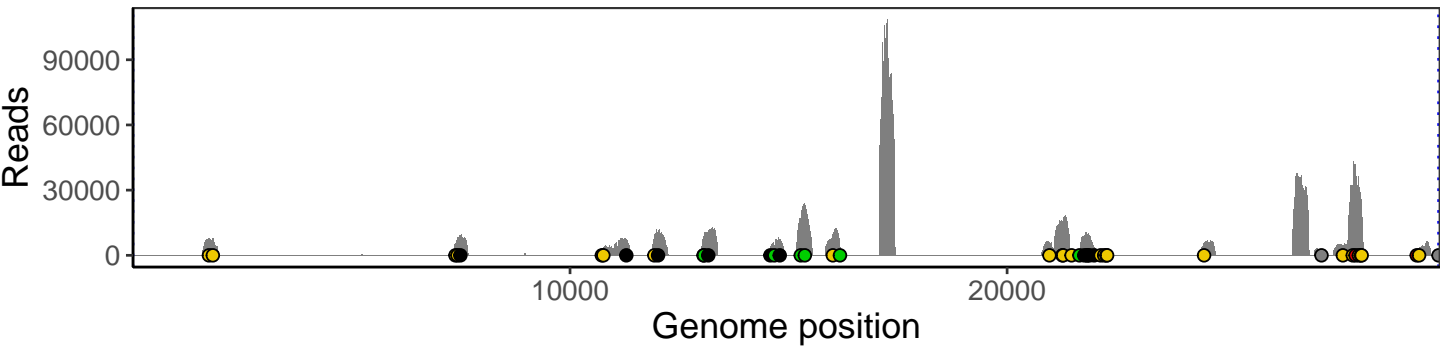


VSP1077-1

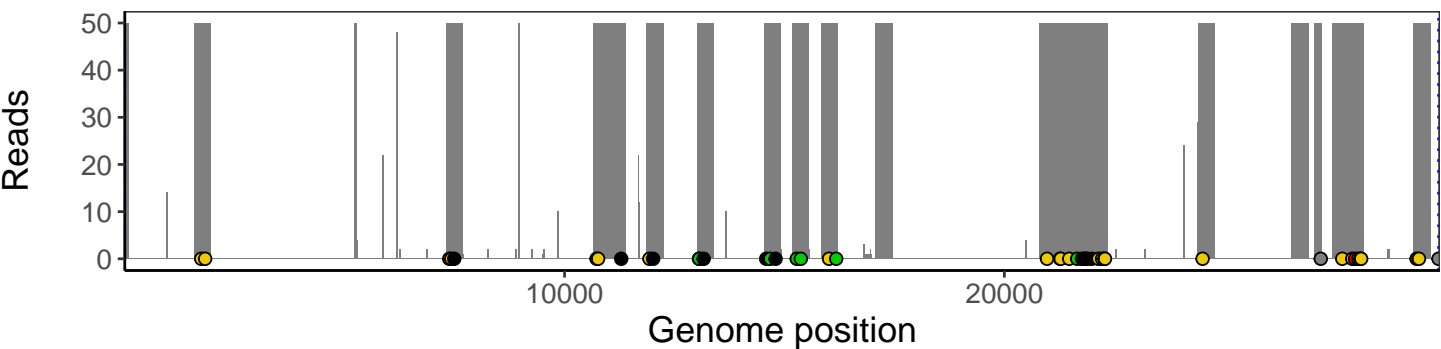
# Analyses of individual experiments and composite results

VSP1077-1 | 2021-03-15 | Saline | HUP Q-0045 | genomes | single experiment

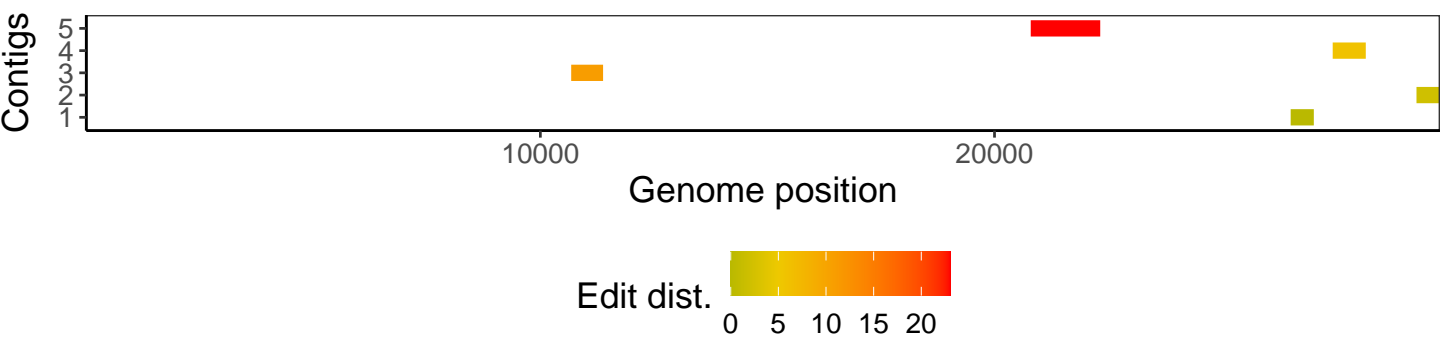
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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