

# COVID-19 subject HUP Q-0195

*2021-05-05*

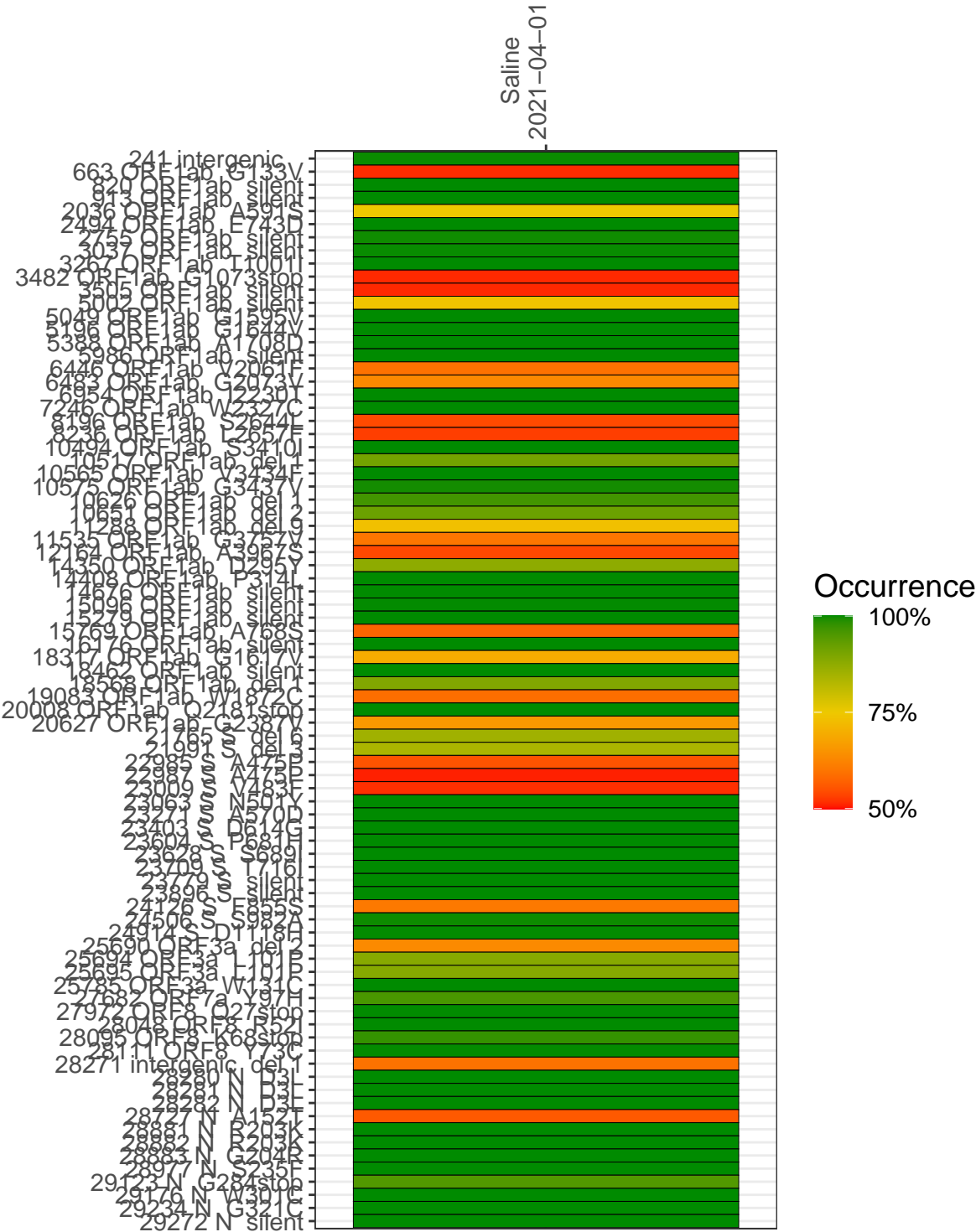
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1758-1	single experiment	NA	Saline	2021-04-01	24.63	B.1.1.7	96.8%	96.8%

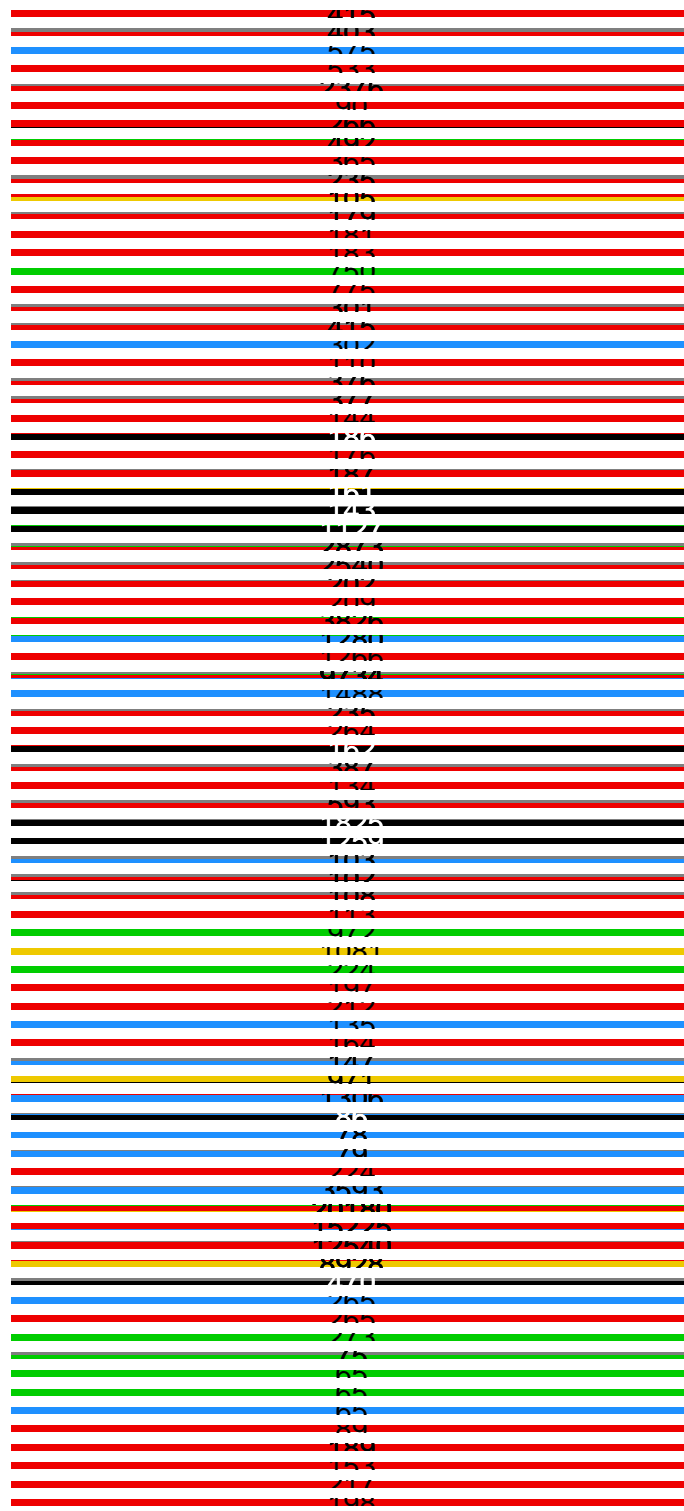
**Variants shared across samples**

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  
2021-04-01

241 INTERGENIC  
653 CURET20 14133V  
820 CURET20 SILENT  
913 CURET20 SILENT  
2036 CURET20 Δ501S  
2494 CURET20 F743T  
2722 CURET20 SILENT  
3137 CURET20 SILENT  
3267 CURET20 11010T  
3482 CURET20 121073CTON  
3505 CURET20 SILENT  
5002 CURET20 SILENT  
5049 CURET20 141505V  
5196 CURET20 141644V  
5388 CURET20 Δ1701AT  
5986 CURET20 SILENT  
6446 CURET20 V2063T  
6783 CURET20 142013V  
6954 CURET20 12230T  
7246 CURET20 W7437T  
8106 CURET20 S2644T  
8236 CURET20 120637C  
10494 CURET20 S3410T  
10517 CURET20 Δ61T  
10565 CURET20 V3434F  
10575 CURET20 143737V  
10626 CURET20 Δ61T  
10651 CURET20 Δ61T  
11288 CURET20 Δ61T  
11535 CURET20 143737V  
12164 CURET20 Δ3967S  
14350 CURET20 11296V  
14408 CURET20 F314T  
14676 CURET20 SILENT  
15096 CURET20 SILENT  
15279 CURET20 SILENT  
15769 CURET20 Δ768S  
16176 CURET20 SILENT  
18317 CURET20 141617V  
18462 CURET20 SILENT  
18568 CURET20 Δ61T  
19083 CURET20 W1877T  
20008 CURET20 112181CTON  
20627 CURET20 142387V  
21765 S Δ61T  
21991 S Δ61T  
22085 S Δ475P  
22487 S Δ475P  
23009 S V483F  
23063 S N401Y  
23271 S Δ5700T  
23403 S 10814T  
23604 S P681H  
23628 S S689T  
23709 S 1716T  
23779 S SILENT  
23896 S SILENT  
24126 S F855S  
24506 S S987Δ  
24914 S 111118H  
25600 CURET20 Δ61T  
25694 CURET20 11010T  
25695 CURET20 11010T  
25785 CURET20 W131T  
27682 CURET20 Y97H  
27972 CURET20 11275CTON  
28048 CURET20 R52T  
28085 CURET20 K682CTON  
28111 CURET20 Y730T  
28271 INTERGENIC Δ61T  
28280 IN 113T  
28281 IN 113T  
28282 IN 113T  
28277 IN Δ152T  
28881 IN R203K  
28882 IN R203K  
28883 IN 14204K  
28977 IN S235F  
29123 IN 14281CTON  
29176 IN W3011T  
29234 IN 14321T  
29272 IN SILENT



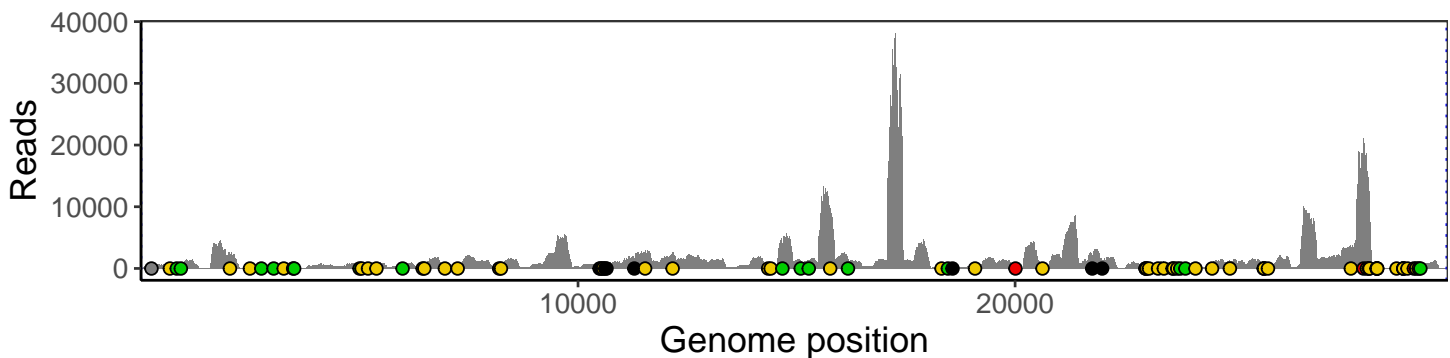
Base change

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

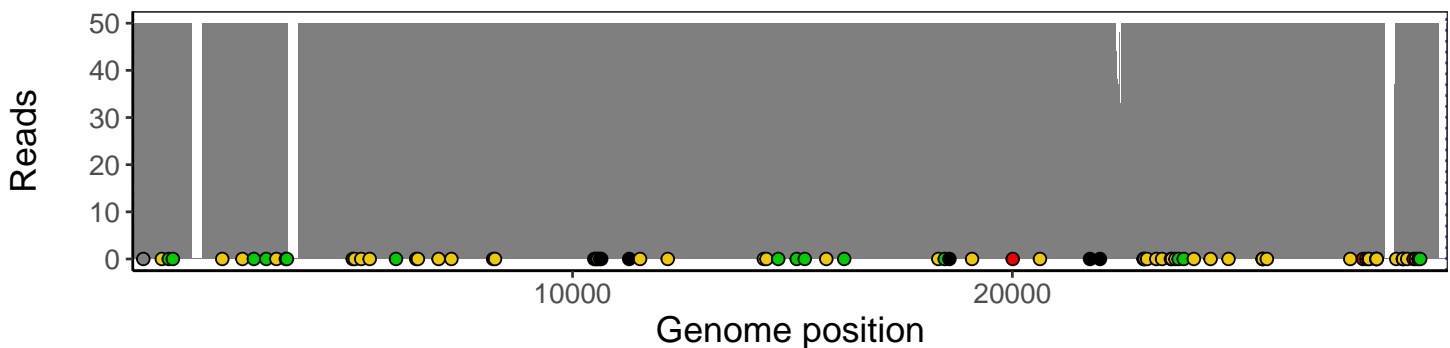
## Analyses of individual experiments and composite results

VSP1758-1 | 2021-04-01 | Saline | HUP Q-0195 | genomes | single experiment

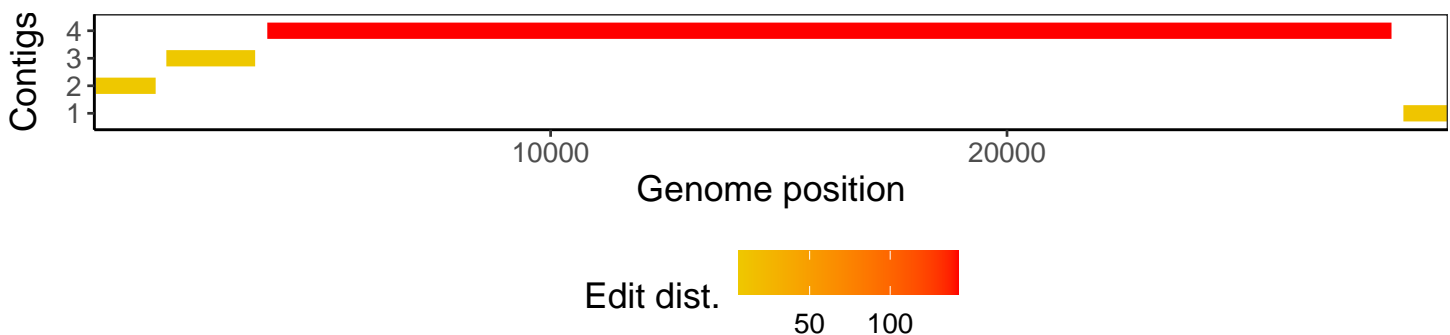
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	3.3.3
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