

# COVID-19 subject UPHS-0099

*2021-06-23*

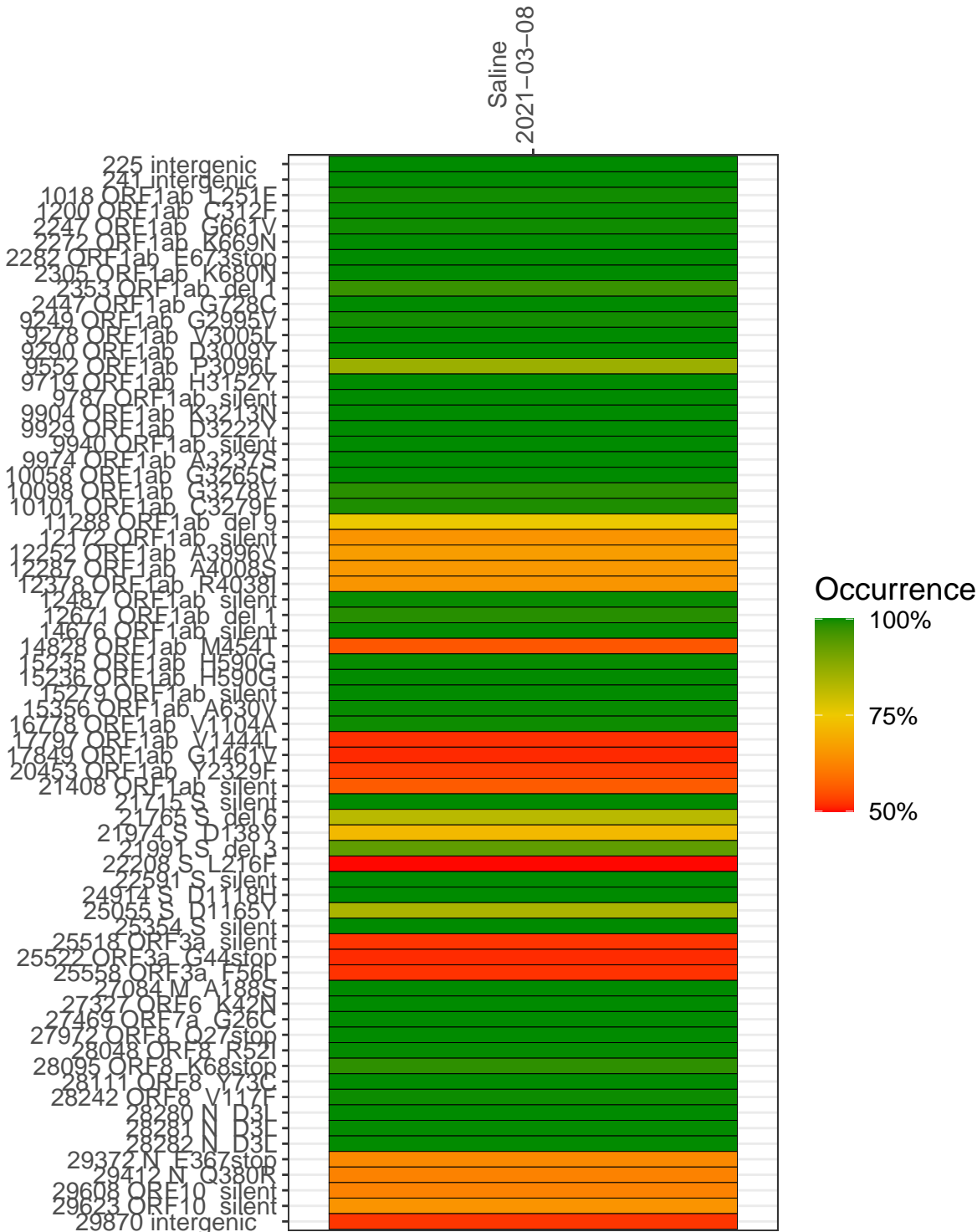
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1030-1	single experiment	NA	Saline	2021-03-08	1.25	NA	38.3%	37.0%

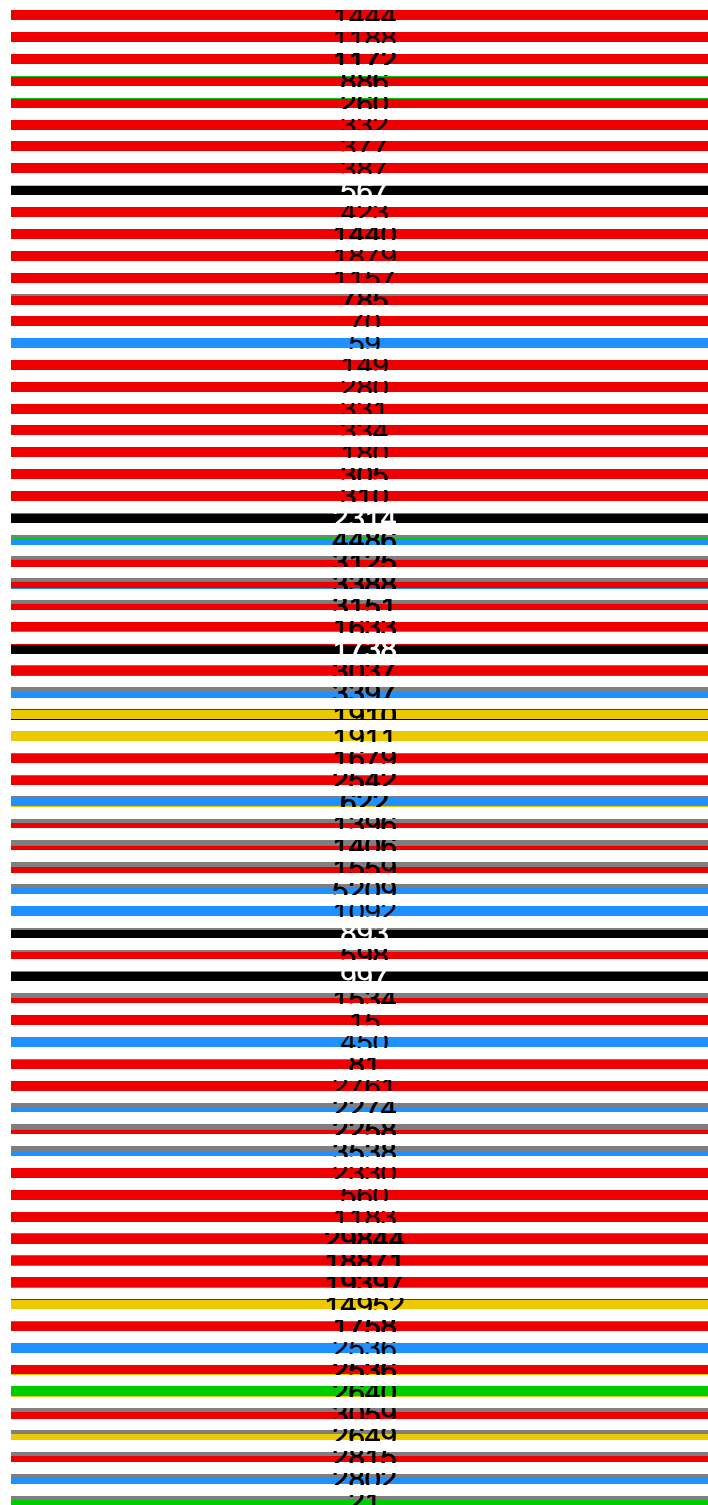
## Variants shared across samples

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  
2021-03-08

225 intergenic  
241 intergenic  
10118 ORF1ab T251F  
12000 ORF1ab C317F  
2247 ORF1ab G661V  
2272 ORF1ab K669N  
2282 ORF1ab F673stop  
2305 ORF1ab K680N  
2353 ORF1ab del 1  
2447 ORF1ab G728C  
9249 ORF1ab G7995V  
9278 ORF1ab V3005I  
9290 ORF1ab D3009Y  
9552 ORF1ab P3096I  
9719 ORF1ab H3152Y  
9787 ORF1ab silent  
9904 ORF1ab K3213N  
9929 ORF1ab D3222Y  
9940 ORF1ab silent  
9974 ORF1ab A3237S  
10058 ORF1ab G3265C  
10098 ORF1ab G3278V  
10101 ORF1ab C3279F  
11288 ORF1ab del 9  
12172 ORF1ab silent  
12252 ORF1ab A3996V  
12287 ORF1ab A4008S  
12378 ORF1ab R4038I  
12487 ORF1ab silent  
12671 ORF1ab del 1  
14676 ORF1ab silent  
14828 ORF1ab M454I  
15235 ORF1ab H5900G  
15236 ORF1ab H5900G  
15279 ORF1ab silent  
15356 ORF1ab A630V  
16778 ORF1ab V1104A  
17797 ORF1ab V1444I  
17849 ORF1ab G1461V  
20453 ORF1ab Y2329F  
21408 ORF1ab silent  
21715 S silent  
21765 S del 6  
21974 S D138Y  
21991 S del 3  
22208 S T216F  
22591 S silent  
24914 S D1118H  
25055 S D1165Y  
25354 S silent  
25518 ORF3a silent  
25522 ORF3a G44stop  
25558 ORF3a F56I  
27084 M A188S  
27327 ORF6 K42N  
27469 ORF7a G26C  
27972 ORF8 D27stop  
28048 ORF8 R52I  
28095 ORF8 K68stop  
28111 ORF8 Y73C  
28242 ORF8 V117F  
28280 N D3I  
28281 N D3I  
28282 N D3I  
29372 N F367stop  
29412 N D380R  
29608 ORF10 silent  
29623 ORF10 silent  
29870 intergenic



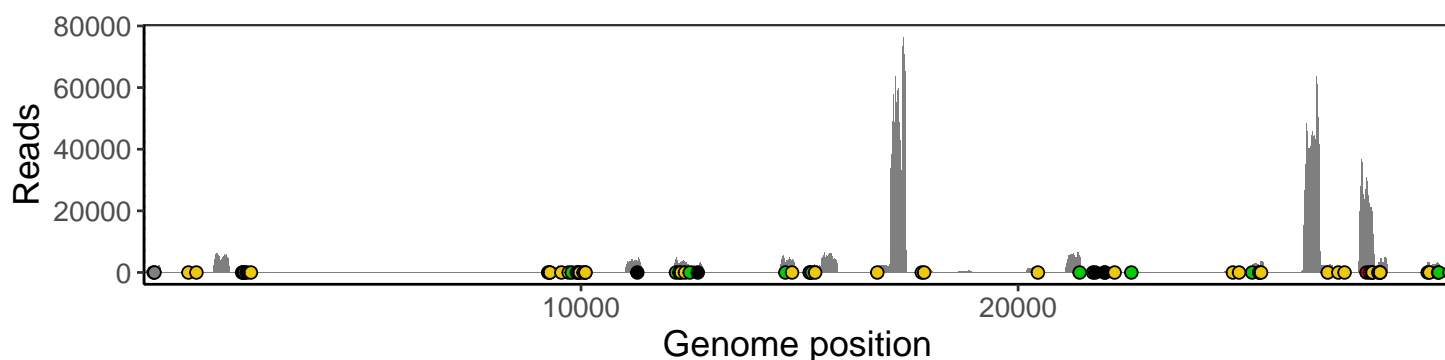
Base change



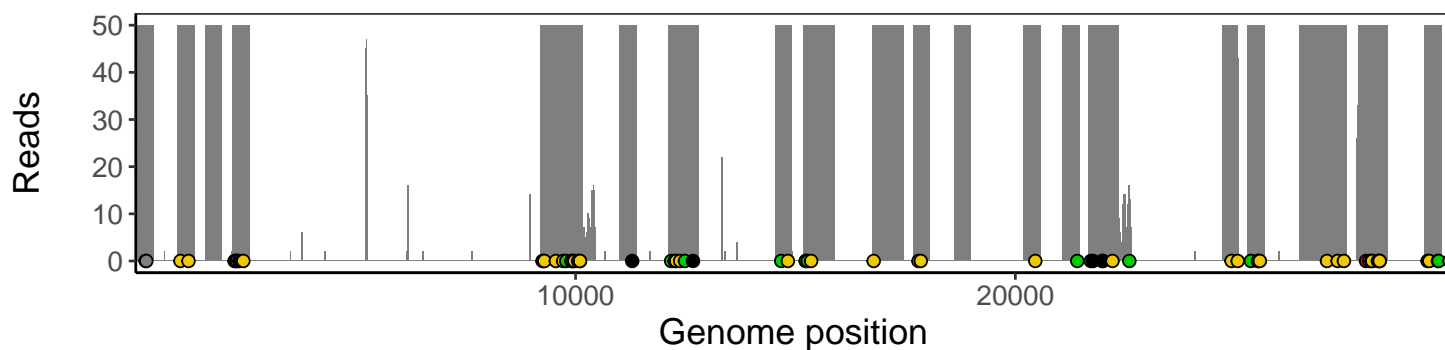
## Analyses of individual experiments and composite results

VSP1030-1 | 2021-03-08 | Saline | UPHS-0099 | genomes | single experiment

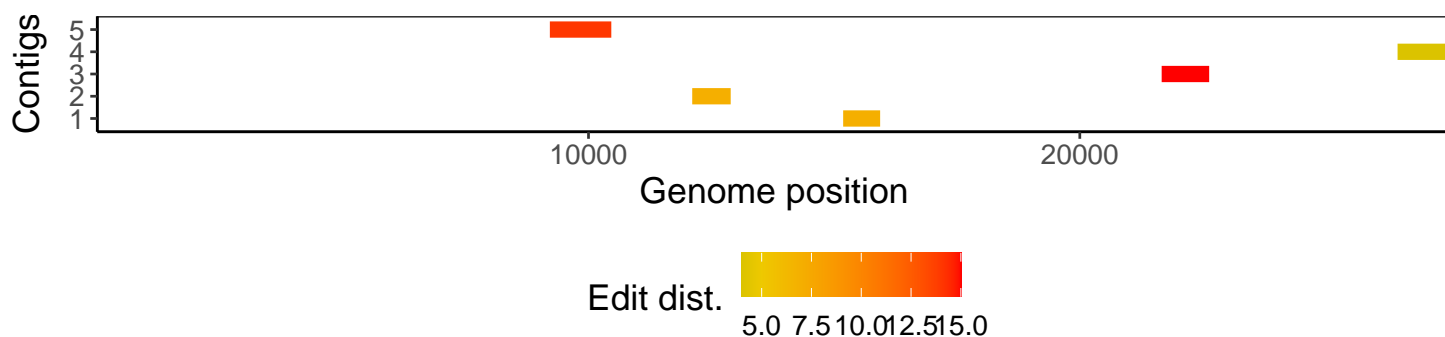
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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