

# COVID-19 subject UPHS-0044

*2021-05-05*

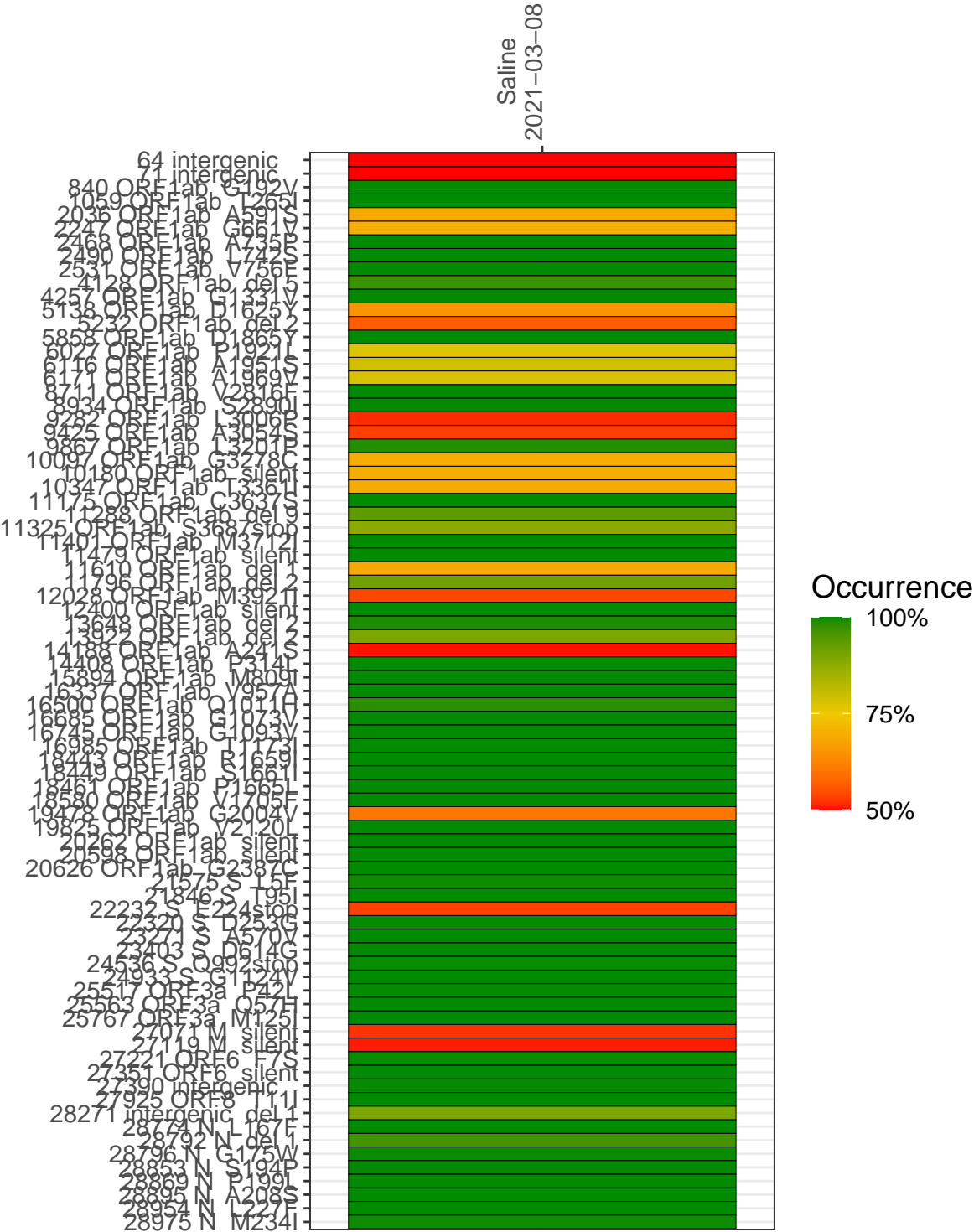
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP0976-1	single experiment	NA	Saline	2021-03-08	3.86	NA	74.3%	73.6%

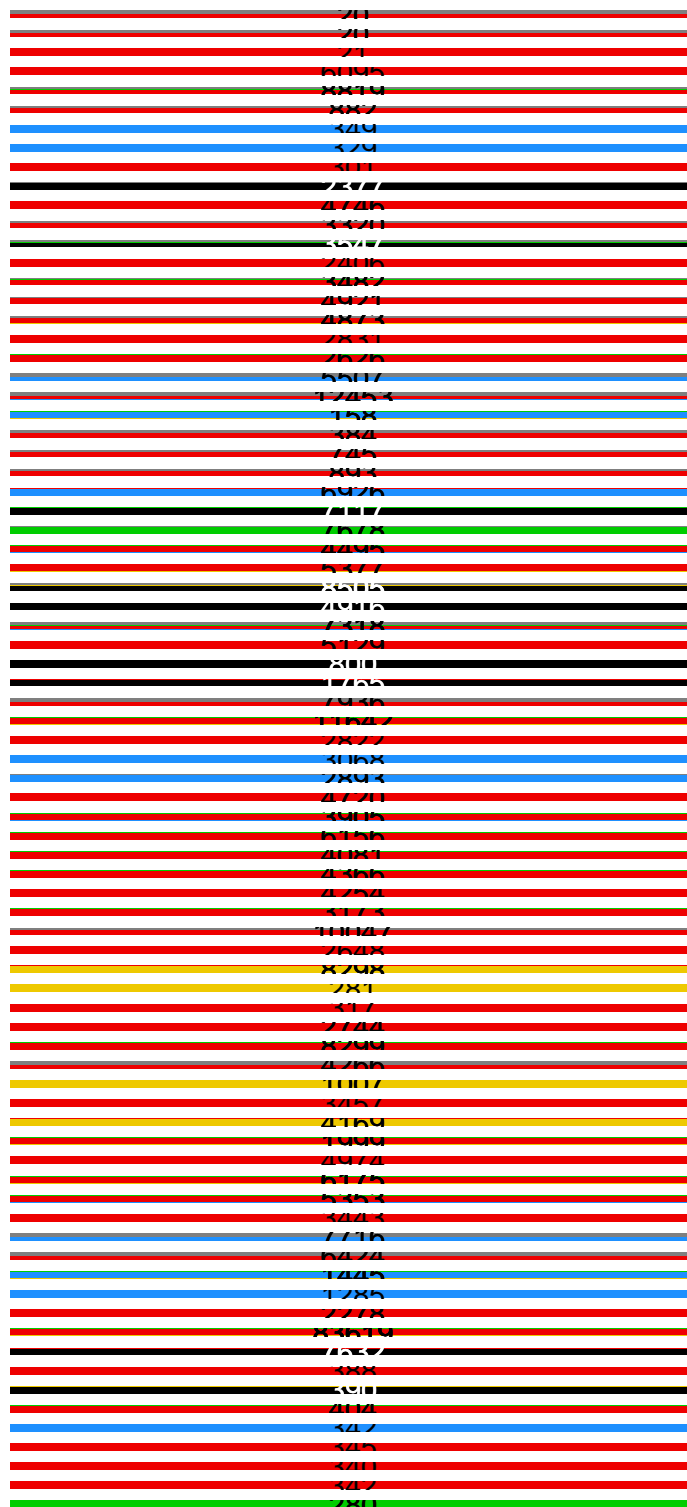
**Variants shared across samples**

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  
2021-03-08

h4 intergenic  
71 intergenic  
x4011 CURE120 14192V  
11154 CURE120 17651  
2036 CURE120 A591S  
2247 CURE120 15661V  
2468 CURE120 A735P  
2490 CURE120 1742S  
2531 CURE120 V756P  
4178 CURE120 del5  
4257 CURE120 141331V  
5138 CURE120 111625Y  
5232 CURE120 del2  
5858 CURE120 111865Y  
6027 CURE120 P19271  
6116 CURE120 A1951S  
6171 CURE120 A1969V  
8711 CURE120 V2816A  
8934 CURE120 S2890I  
9282 CURE120 13006P  
9425 CURE120 A3054S  
9867 CURE120 13201P  
10097 CURE120 143278I  
10180 CURE120 silent  
10142 CURE120 13361I  
11175 CURE120 13637S  
11288 CURE120 del9  
11325 CURE120 S36870N  
11401 CURE120 M3712I  
11479 CURE120 silent  
11611 CURE120 del1  
11796 CURE120 del2  
12028 CURE120 M3921I  
12400 CURE120 silent  
13648 CURE120 del2  
13972 CURE120 del2  
14188 CURE120 A241S  
14408 CURE120 P314I  
15894 CURE120 M1809I  
16337 CURE120 V957A  
16500 CURE120 111017H  
16685 CURE120 141073V  
16745 CURE120 141093V  
16985 CURE120 11173I  
18443 CURE120 R1659I  
18449 CURE120 S1661I  
18461 CURE120 P1660I  
18580 CURE120 V1701P  
19478 CURE120 142004V  
19825 CURE120 V2120I  
20262 CURE120 silent  
20598 CURE120 silent  
20626 CURE120 142387C  
21575 S15E  
21846 S19E  
22332 S F2240N  
22320 S 112231S  
23271 S A570V  
23403 S 10614I  
24536 S 109920N  
24933 S 141124V  
25517 CURE32 P42I  
25563 CURE32 1057H  
25767 CURE32 M125I  
27071 M silent  
27119 M silent  
27271 CURE32 F7S  
27351 CURE32 silent  
27491 intergenic  
27925 CURE32 111  
28271 intergenic del1  
28774 M 1767E  
28792 M del1  
28796 M 14175W  
28833 M S194P  
28869 M P199I  
28895 M A708S  
28954 M 1727E  
28975 M M234I

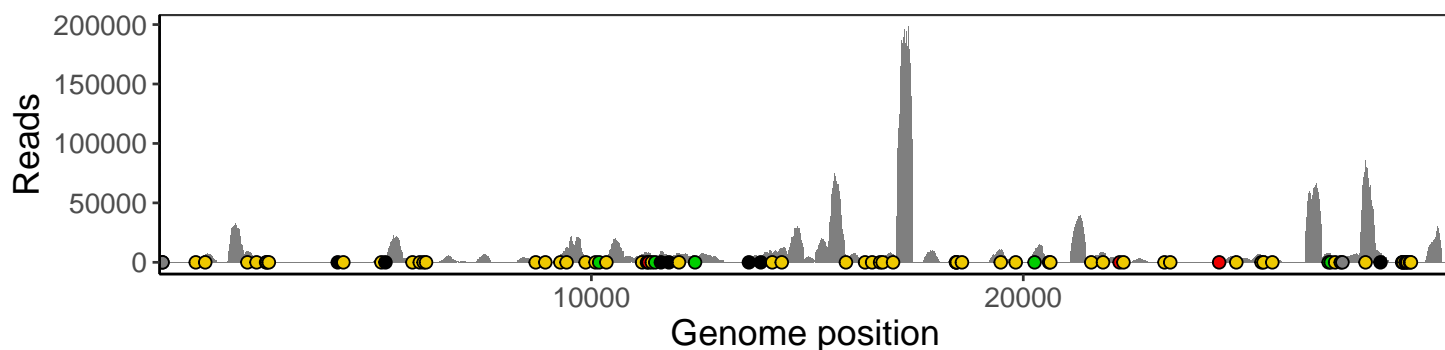


Base change  
Expected  
A  
T  
C  
G  
N  
Ins/Del  
No data

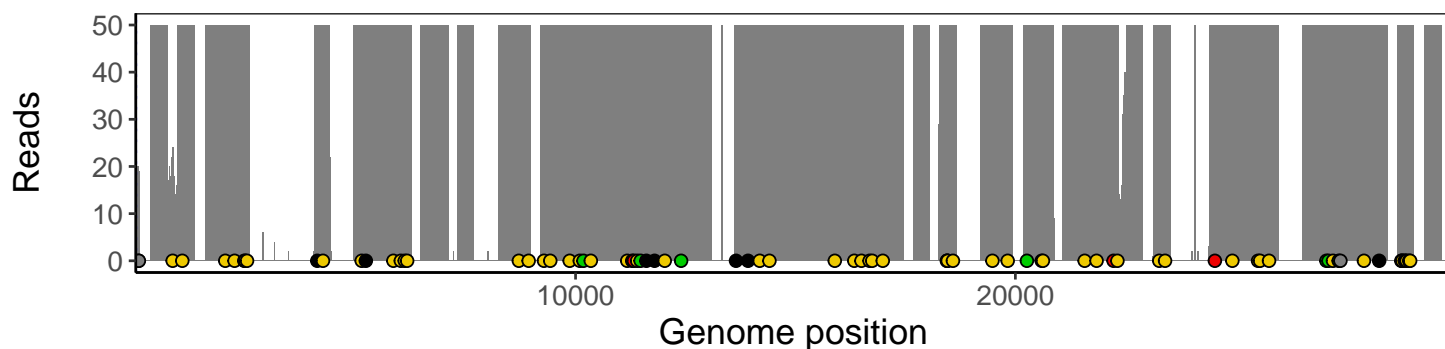
## Analyses of individual experiments and composite results

VSP0976-1 | 2021-03-08 | Saline | UPHS-0044 | 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	3.0.0
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