

COVID-19 subject UPHS-0026

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.

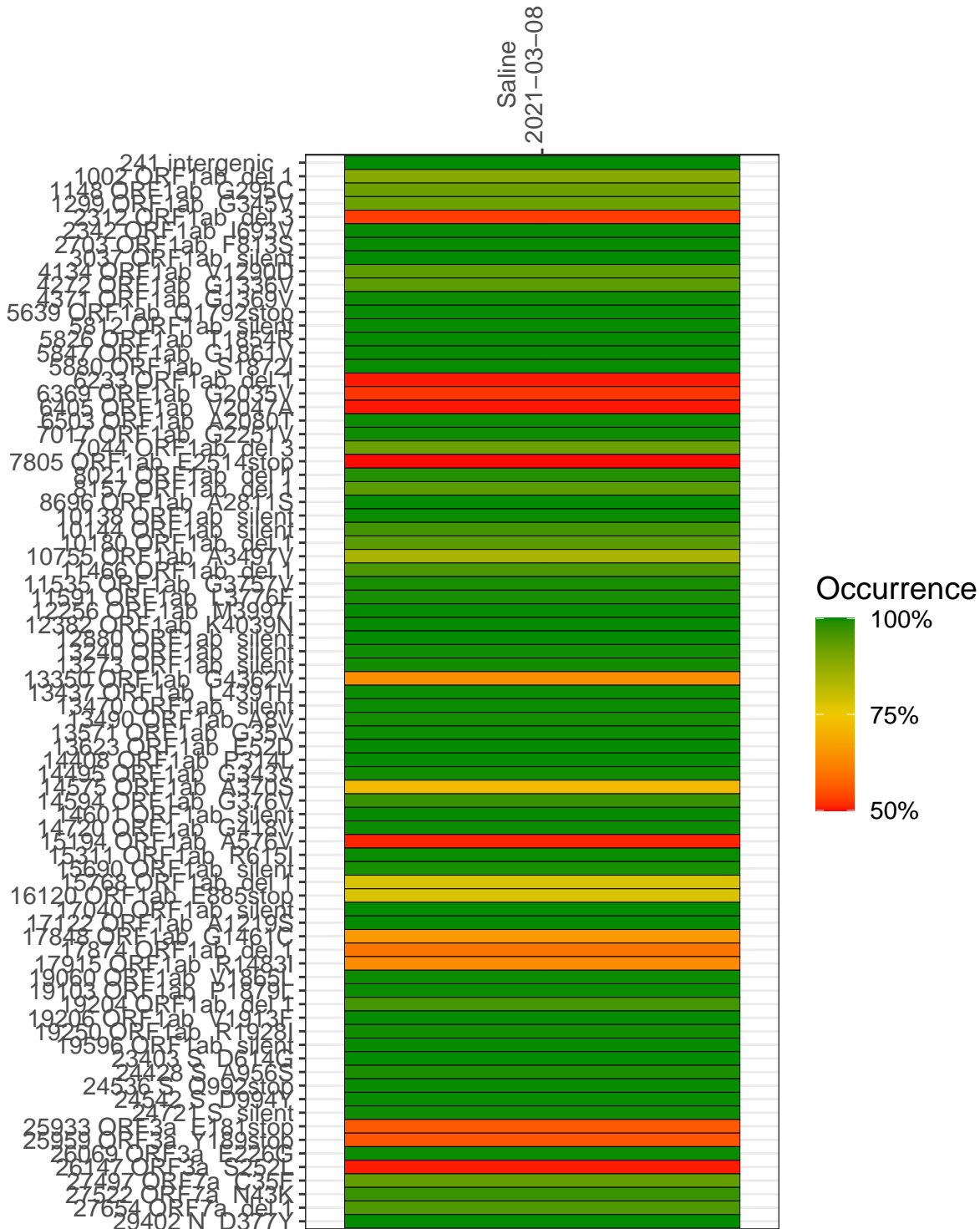
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with $> 90\%$ sequence coverage.

Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (≥ 5 reads)
VSP0958-1	single experiment	NA	Saline	2021-03-08	8.44	NA	97.9%	92.1%

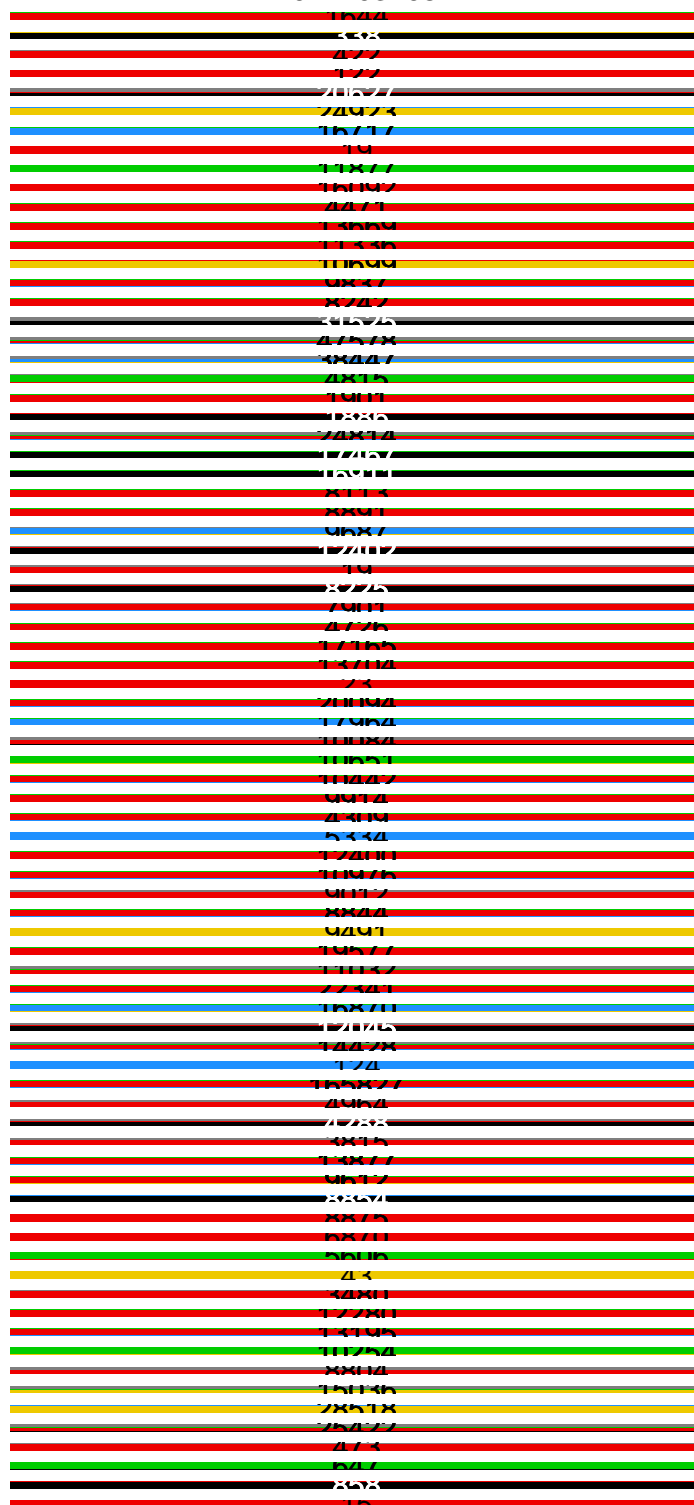
Variants shared across samples

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

241 intergenic
11112 URF120 DEL T
11148 URF120 C2295C
11299 URF120 C3345V
12312 URF120 DEL C
12342 URF120 H693V
12713 URF120 F413S
13137 URF120 SHUNT
13134 URF120 V1290I
12772 URF120 C41336V
14371 URF120 C41369V
15339 URF120 C11792STON
15412 URF120 SHUNT
15426 URF120 C11854R
15447 URF120 C41861V
15481 URF120 S1872I
16233 URF120 DEL T
16369 URF120 C42013V
16405 URF120 V21047A
16503 URF120 A21080I
17017 URF120 C42251V
17044 URF120 DEL C
17805 URF120 F2514STON
18121 URF120 DEL T
18157 URF120 DEL T
18496 URF120 A2811S
19113 URF120 SHUNT
19144 URF120 SHUNT
19180 URF120 DEL T
19195 URF120 A3297V
19166 URF120 DEL T
19153 URF120 C43757V
19191 URF120 T3776F
19256 URF120 M3997I
19382 URF120 K40190N
19380 URF120 SHUNT
19740 URF120 SHUNT
19773 URF120 SHUNT
19790 URF120 C44367V
19837 URF120 C4491H
19840 URF120 SHUNT
19890 URF120 A48V
19571 URF120 C435V
19623 URF120 F421I
19608 URF120 P314I
19695 URF120 C4343V
19575 URF120 A370S
19594 URF120 C4376V
19601 URF120 SHUNT
19720 URF120 C4418V
19794 URF120 A576V
19311 URF120 R616I
19690 URF120 SHUNT
19768 URF120 DEL T
19720 URF120 F4455STON
19740 URF120 SHUNT
19722 URF120 A1219S
19848 URF120 C4461C
19874 URF120 DEL T
19915 URF120 R1483I
19960 URF120 V1865I
19913 URF120 P1879I
19904 URF120 DEL T
19916 URF120 V1913F
19901 URF120 R1928I
19996 URF120 SHUNT
19403 S 19614C
19428 S 1966S
19536 S 19992STON
19542 S 1994Y
19721 S SHUNT
19933 URF32 F181STON
19954 URF32 Y184STON
19964 URF32 F2261C
19747 URF32 S252I
19749 URF32 C335F
19752 URF32 N438K
19754 URF32 DEL T
194012 N 11377Y



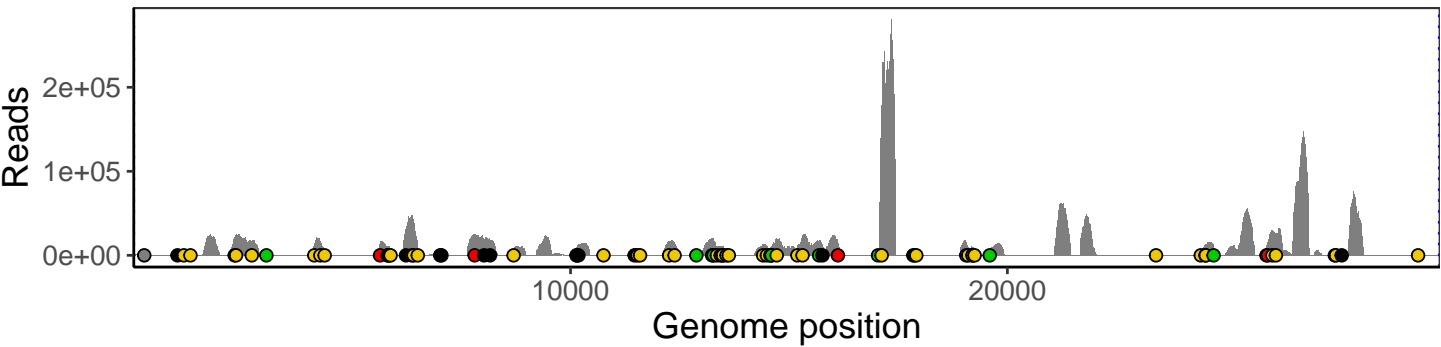
Base change

Expected
A
T
C
G
N
Ins/Del
No data

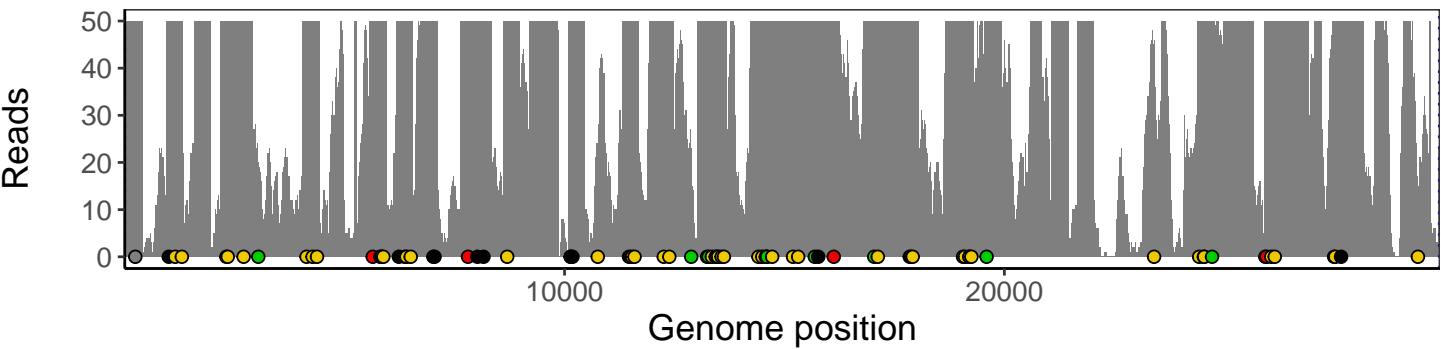
Analyses of individual experiments and composite results

VSP0958-1 | 2021-03-08 | Saline | UPHS-0026 | genomes | single experiment

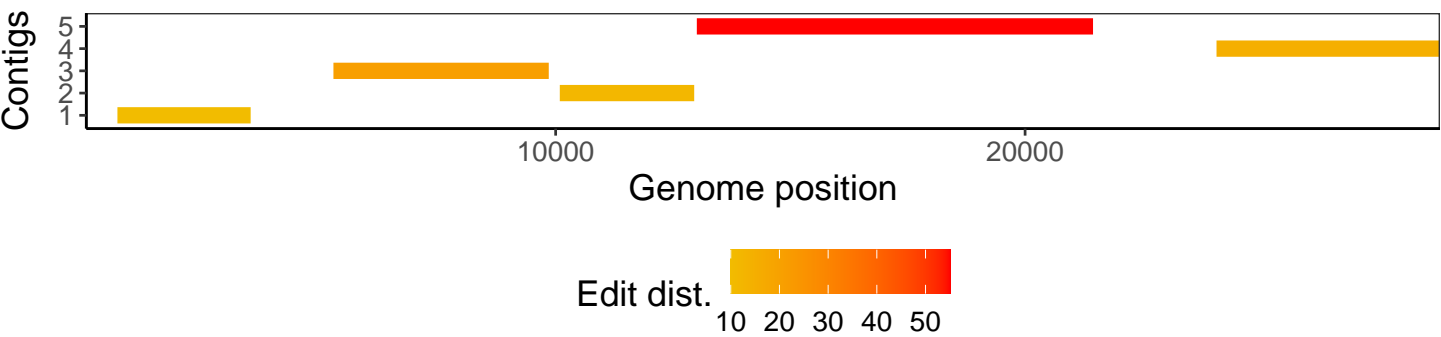
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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