

COVID-19 subject UPHS-0026

2021-05-05

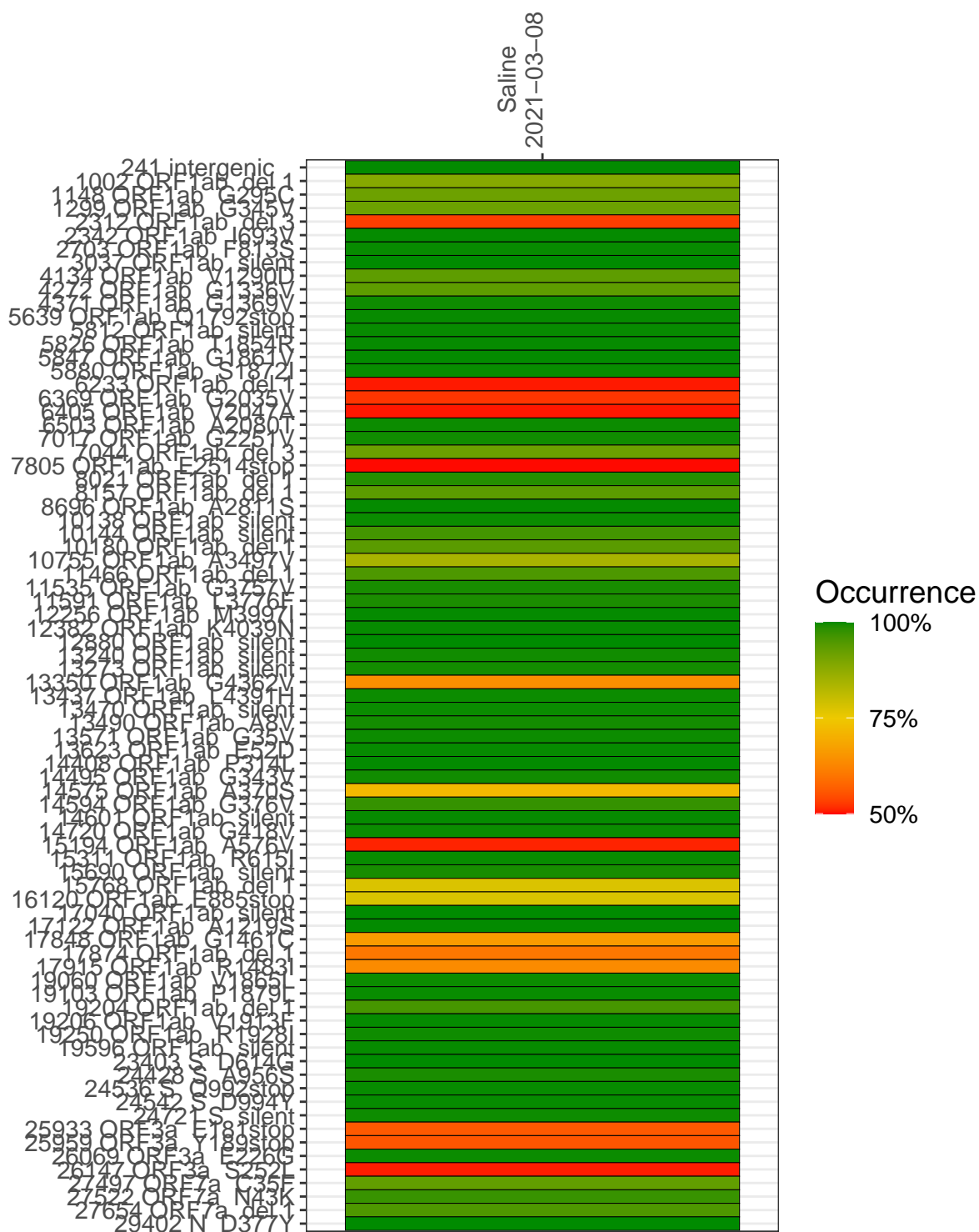
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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 (≥ 5 reads)
VSP0958-1	single experiment	NA	Saline	2021-03-08	8.44	NA	97.9%	92.1%

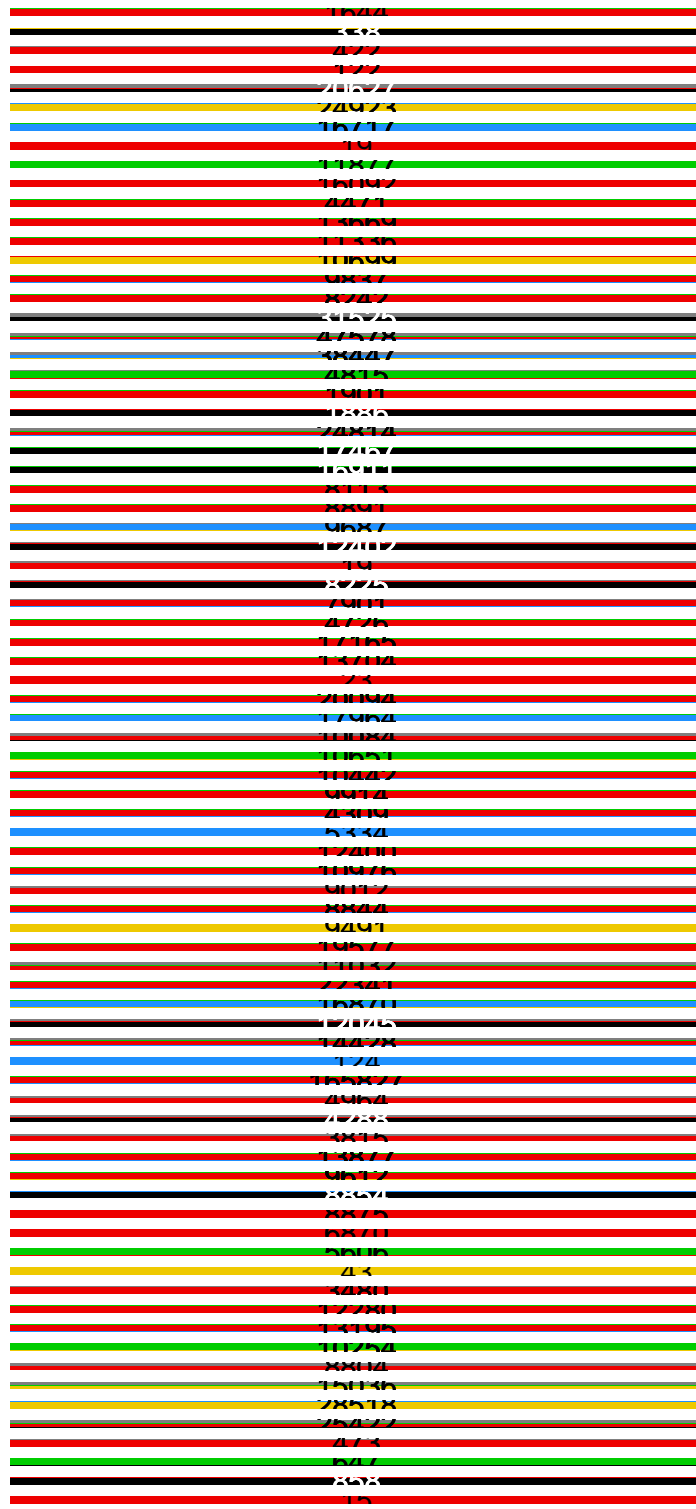
Variants shared across samples

The heat map below shows how variants (reference genome `/home/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

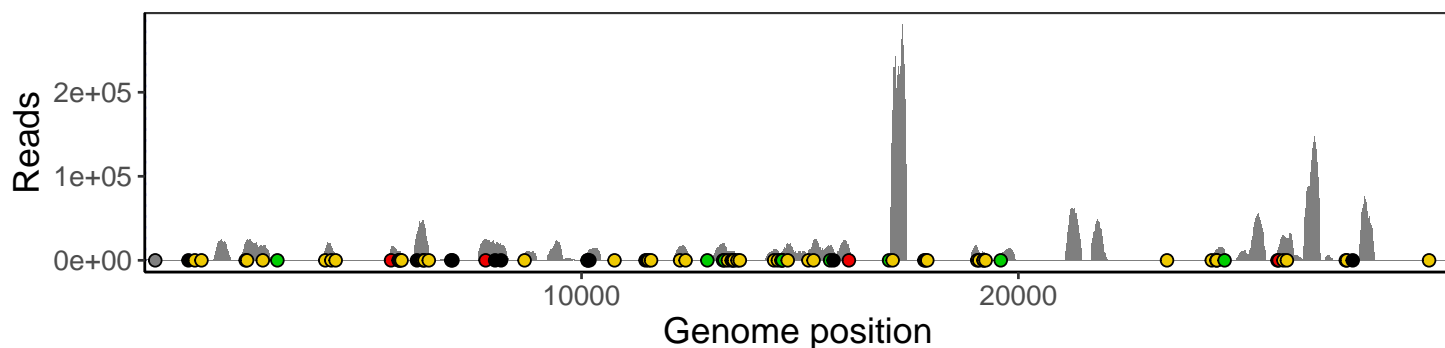
241 intergenic
11012 CURE120 DEL T
11148 CURE120 C2295C
11299 CURE120 C3345V
11312 CURE120 DEL C
11342 CURE120 1693V
11113 CURE120 F413S
10137 CURE120 SHUNT
11134 CURE120 V1290U
12772 CURE120 C41336V
14371 CURE120 C41369V
16394 CURE120 C11792STON
18112 CURE120 SHUNT
18276 CURE120 C11854R
18417 CURE120 C41861V
18480 CURE120 S1872I
18733 CURE120 DEL T
18694 CURE120 C420136V
16105 CURE120 V121047A
16103 CURE120 A21080U
10117 CURE120 C42251V
11144 CURE120 DEL C
18105 CURE120 F2514STON
10121 CURE120 DEL T
18157 CURE120 DEL T
18996 CURE120 A2811S
111138 CURE120 SHUNT
111144 CURE120 SHUNT
111180 CURE120 DEL T
111755 CURE120 A32097V
111466 CURE120 DEL T
111535 CURE120 C43757V
11591 CURE120 C13776R
112256 CURE120 M13997I
112382 CURE120 K41190U
112880 CURE120 SHUNT
113240 CURE120 SHUNT
113773 CURE120 SHUNT
113390 CURE120 C42367V
113437 CURE120 C4391H
113470 CURE120 SHUNT
113490 CURE120 A8V
113571 CURE120 C435V
113623 CURE120 F421I
114018 CURE120 P4314I
114495 CURE120 C4343V
114575 CURE120 A370S
114594 CURE120 C4376V
114611 CURE120 SHUNT
114720 CURE120 C4418V
115194 CURE120 A576V
115311 CURE120 R616I
116901 CURE120 SHUNT
116768 CURE120 DEL T
116120 CURE120 F4455STON
117101 CURE120 SHUNT
117122 CURE120 A1219S
117448 CURE120 C41461C
117874 CURE120 DEL T
117915 CURE120 R1483I
118001 CURE120 V1866I
119113 CURE120 P1879I
119104 CURE120 DEL T
119216 CURE120 V1913F
119201 CURE120 R1928I
119596 CURE120 SHUNT
1134013 S 116141C
114478 S 11956S
114536 S 11992STON
114542 S 11994V
114721 S SHUNT
115933 CURE12 F1818STON
115954 CURE12 Y1818STON
116164 CURE12 F2261C
116147 CURE12 S252I
117497 CURE12 C135F
117522 CURE12 1843K
117654 CURE12 DEL T
1194112 11377Y



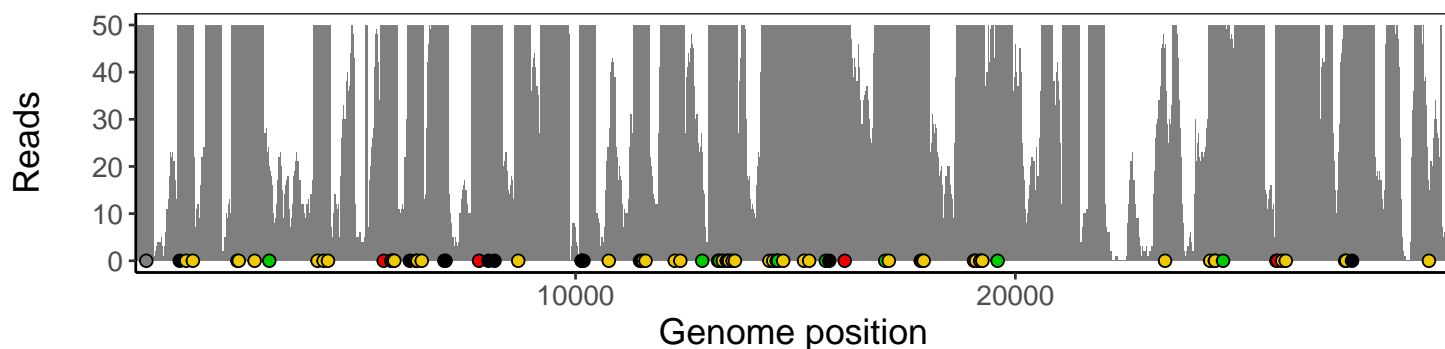
Analyses of individual experiments and composite results

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

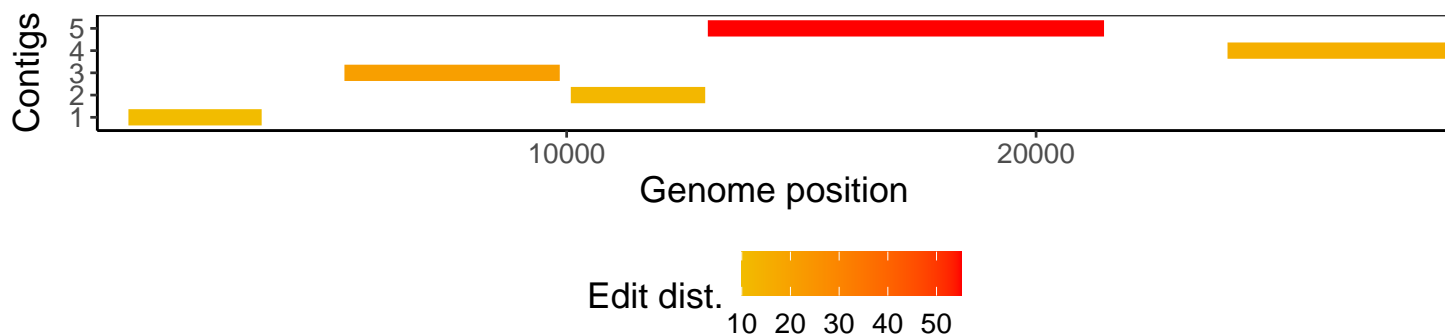
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according to variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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