

COVID-19 subject HUP Q-0190

2021-05-05

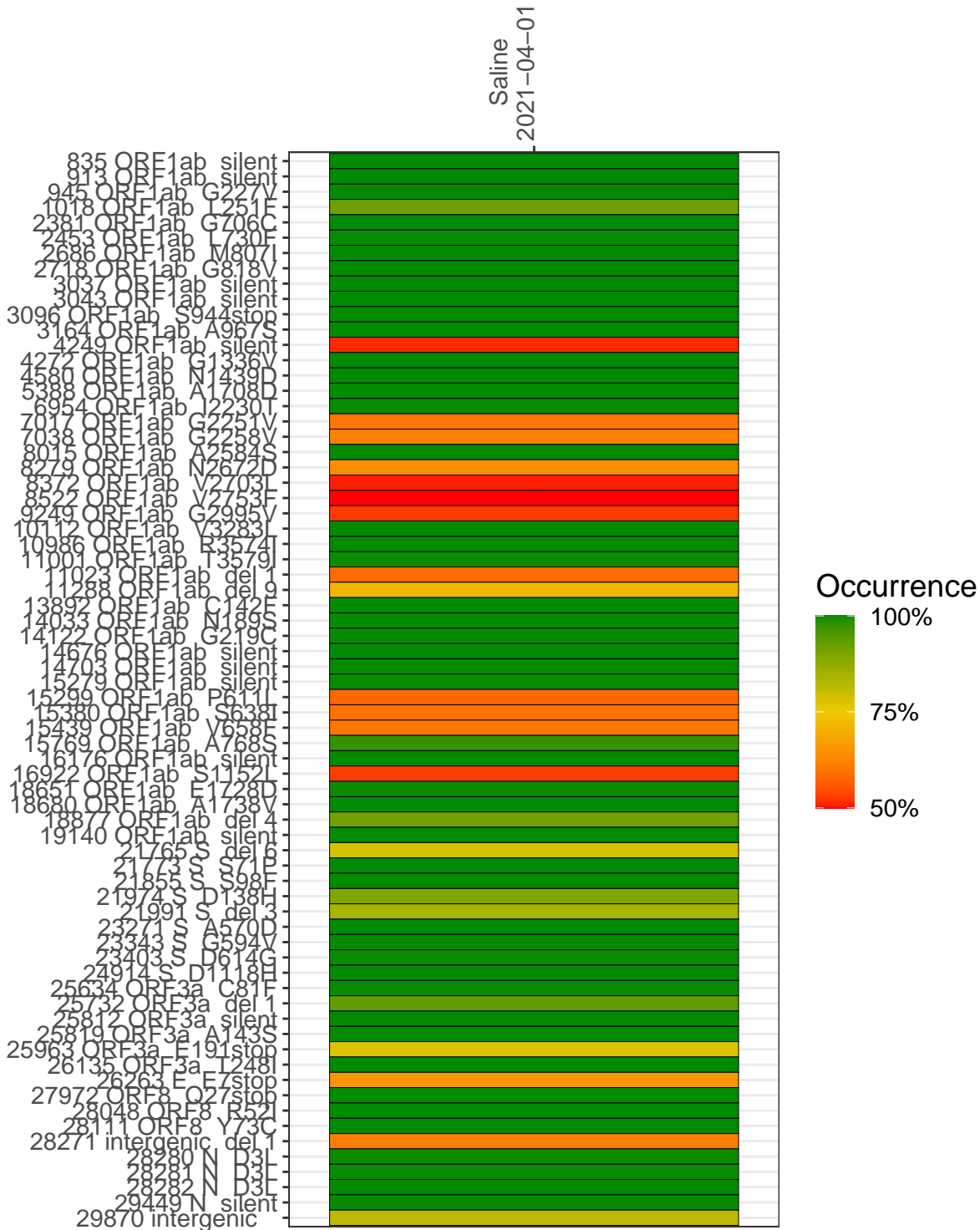
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1753-1	single experiment	NA	Saline	2021-04-01	2.86	NA	59.5%	59.1%

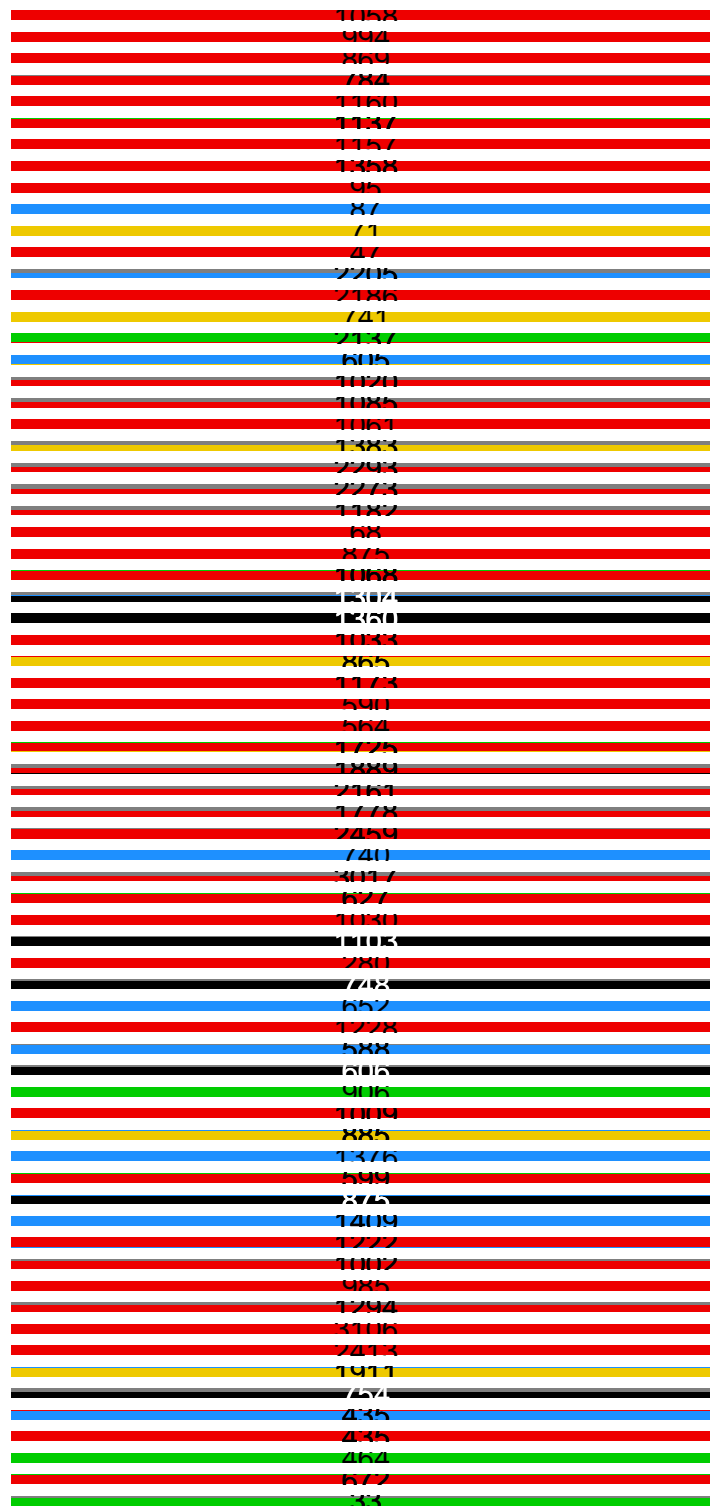
Variants shared across samples

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
2021-04-01

8335 ORF12n silent
9133 ORF12n silent
9425 ORF12n (5227V)
10118 ORF12n T251E
23381 ORF12n (5706E)
24533 ORF12n T730E
26686 ORF12n MX0171
27718 ORF12n (5818V)
30037 ORF12n silent
30043 ORF12n silent
30096 ORF12n S9448ton
31664 ORF12n A967S
42249 ORF12n silent
42772 ORF12n (51336V)
45800 ORF12n N1439U
53888 ORF12n A1708U
69564 ORF12n T2230U
70117 ORF12n (52251V)
70388 ORF12n (52258V)
80075 ORF12n A2584S
82779 ORF12n N2672U
83372 ORF12n V2703E
85222 ORF12n V2753E
92249 ORF12n (52995V)
101112 ORF12n V3283E
101986 ORF12n R3574E
110001 ORF12n T3579U
110123 ORF12n del 1
112288 ORF12n del 9
138922 ORF12n (1142E)
140333 ORF12n N189S
141222 ORF12n (5219U)
14676 ORF12n silent
147033 ORF12n silent
152779 ORF12n silent
152999 ORF12n P611U
153380 ORF12n S638U
154339 ORF12n V658E
15769 ORF12n A768S
16176 ORF12n silent
16922 ORF12n S1152U
18651 ORF12n F1728U
18680 ORF12n A1738V
18877 ORF12n del 4
19140 ORF12n silent
21765 S del 6
21773 S S71P
21855 S S98E
21974 S D13XH
21991 S del 3
23271 S A570U
23343 S (5594V)
23403 S D614U
24914 S D111XH
25634 ORF32 (X1E)
25732 ORF32 del 1
25812 ORF32 silent
25819 ORF32 A143S
25963 ORF32 F1918ton
26135 ORF32 T248U
26263 E F78ton
27972 ORF32 (1278ton)
28048 ORF32 R52U
28111 ORF32 Y73U
28271 intergenic del 1
28280 N T3E
28281 N T3E
28282 N T3E
29449 N silent
29870 intergenic



Base change

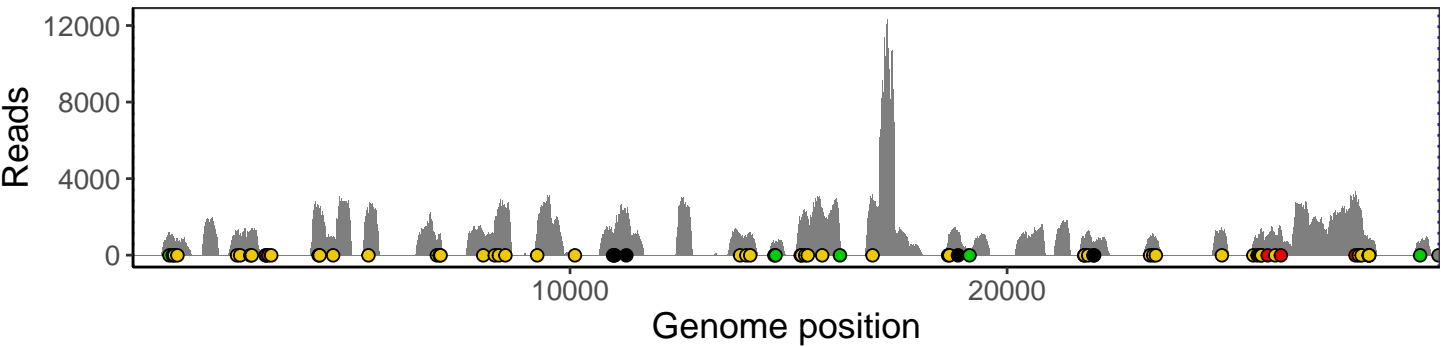
Expected
A
T
C
G
N
Ins/Del
No data

VSP1753-1

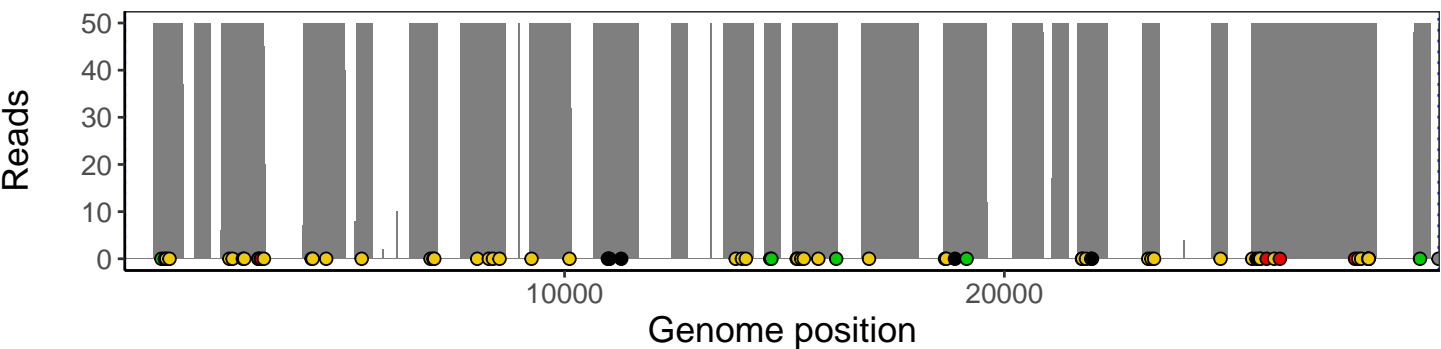
Analyses of individual experiments and composite results

VSP1753-1 | 2021-04-01 | Saline | HUP Q-0190 | genomes | single experiment

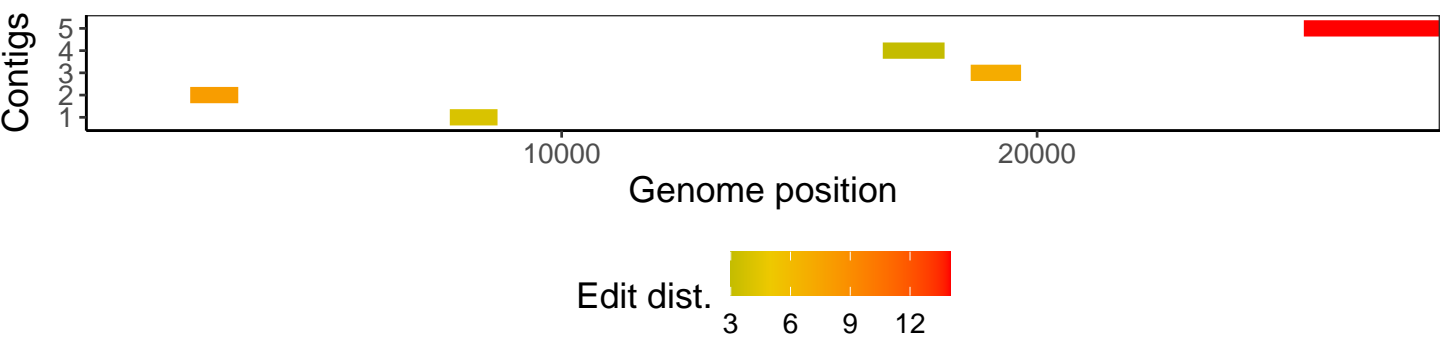
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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