

COVID-19 subject HUP Q-0030

2021-06-23

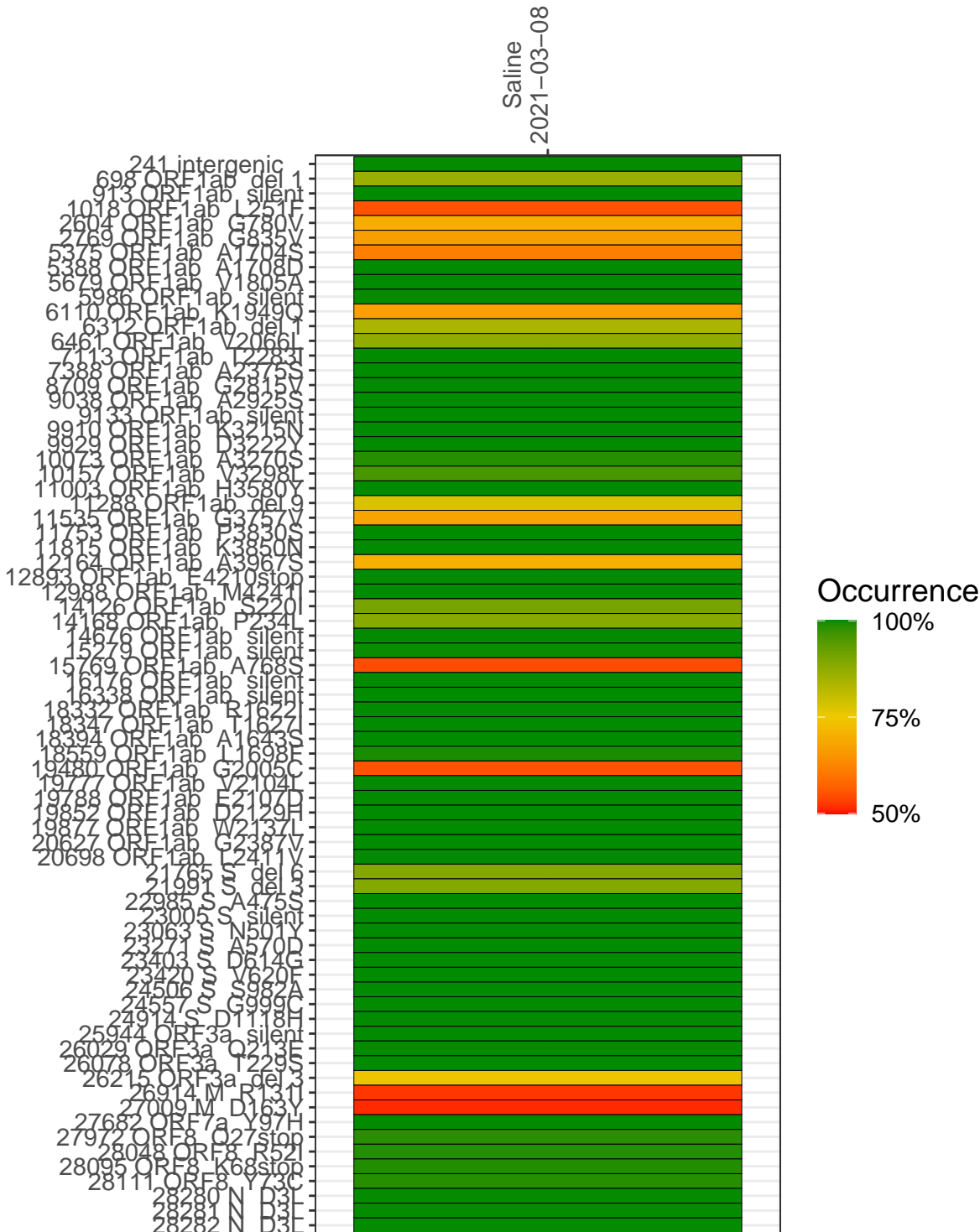
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1032-1	single experiment	NA	Saline	2021-03-08	2.82	NA	74.4%	73.5%

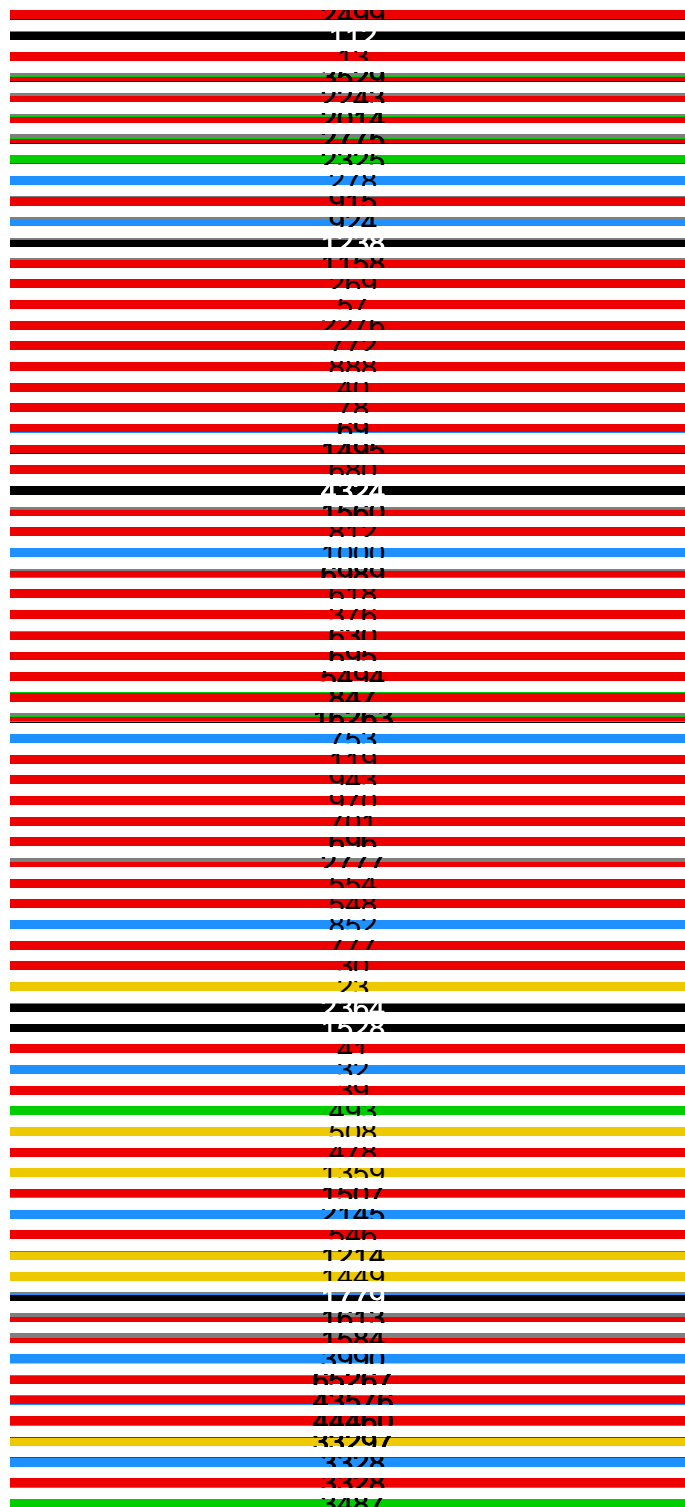
Variants shared across samples

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
698 (ORF12n) del 1
913 (ORF12n) silent
1018 (ORF12n) T251F
2604 (ORF12n) G780V
2769 (ORF12n) G835V
5375 (ORF12n) A1704S
5388 (ORF12n) A1708I
5679 (ORF12n) V1805A
5986 (ORF12n) silent
6110 (ORF12n) K1949Q
6312 (ORF12n) del 1
6461 (ORF12n) V2066I
7113 (ORF12n) T2283I
7388 (ORF12n) A2375S
8709 (ORF12n) G2815V
9038 (ORF12n) A2925S
9133 (ORF12n) silent
9910 (ORF12n) K3215N
9929 (ORF12n) T3222Y
10073 (ORF12n) A3270S
10157 (ORF12n) V3298I
11003 (ORF12n) H3580Y
11288 (ORF12n) del 9
11535 (ORF12n) G3757V
11753 (ORF12n) P3830S
11815 (ORF12n) K3850N
12164 (ORF12n) A3967S
12893 (ORF12n) F4270Istop
12988 (ORF12n) M4241I
14126 (ORF12n) S520I
14168 (ORF12n) P234I
14676 (ORF12n) silent
15279 (ORF12n) silent
15769 (ORF12n) A768S
16176 (ORF12n) silent
16338 (ORF12n) silent
18332 (ORF12n) R1622I
18347 (ORF12n) T1627I
18394 (ORF12n) A1643S
18559 (ORF12n) T1698F
19480 (ORF12n) G2005C
19777 (ORF12n) V2104I
19788 (ORF12n) F2107I
19852 (ORF12n) T2129H
19877 (ORF12n) W2137I
20627 (ORF12n) G2387V
20698 (ORF12n) T2411V
21765 S del 6
21991 S del 3
22985 S A475S
23005 S silent
23063 S N501Y
23271 S A570I
23403 S T614G
23420 S V620F
24506 S S982A
24557 S G999C
24914 S T1118H
25944 (ORF32) silent
26029 (ORF32) C213F
26078 (ORF32) T220S
26215 (ORF32) del 3
26914 M R131I
27009 M T163Y
27682 (ORF72) Y97H
27972 (ORF8) C275stop
28028 (ORF8) R52I
28095 (ORF8) K68stop
28111 (ORF8) Y73C
28280 N T3I
28281 N T3I
28282 N T3I



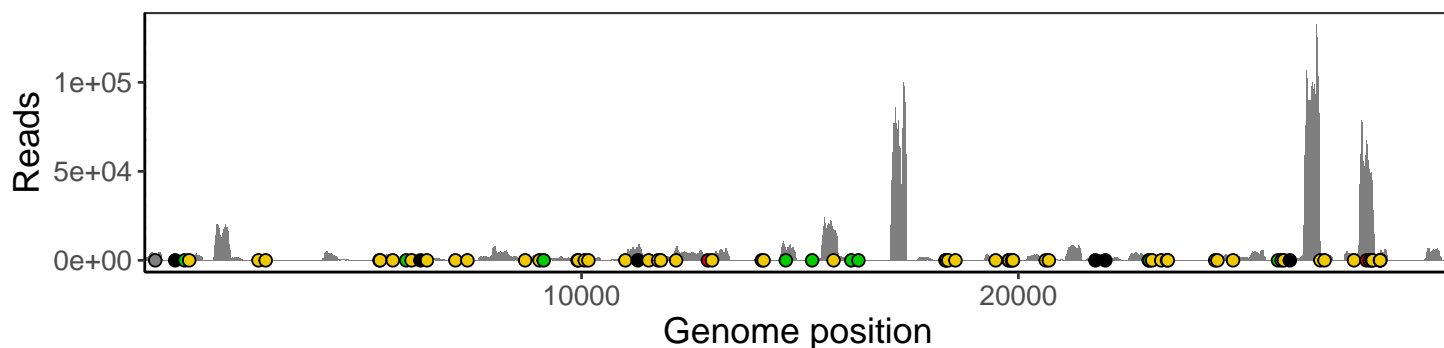
Base change
Expected
A
T
C
G
N
Ins/Del
No data

VSP1032-1

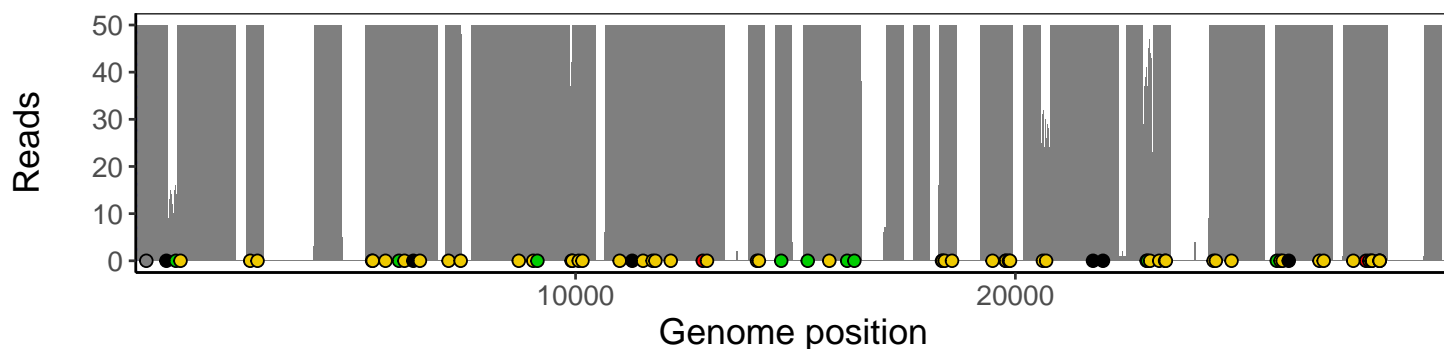
Analyses of individual experiments and composite results

VSP1032-1 | 2021-03-08 | Saline | HUP Q-0030 | genomes | single experiment

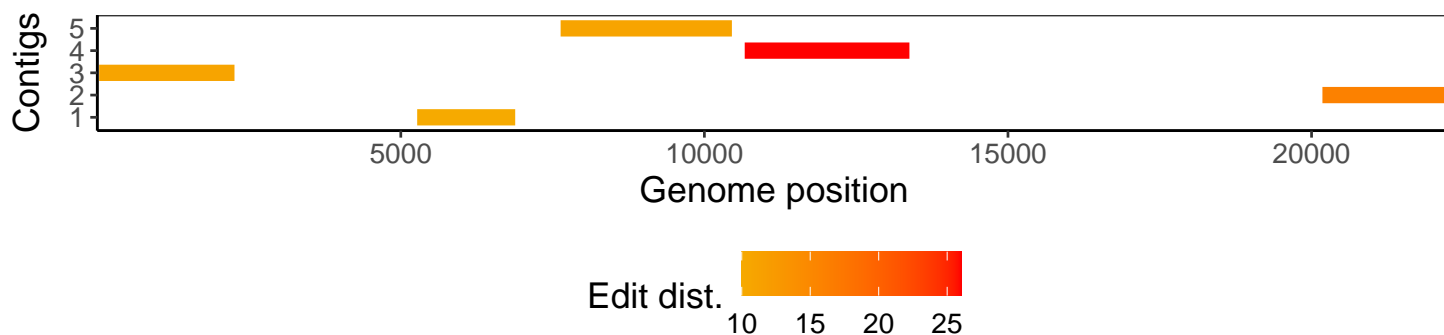
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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