

COVID-19 subject HUP Q-0030

2021-05-05

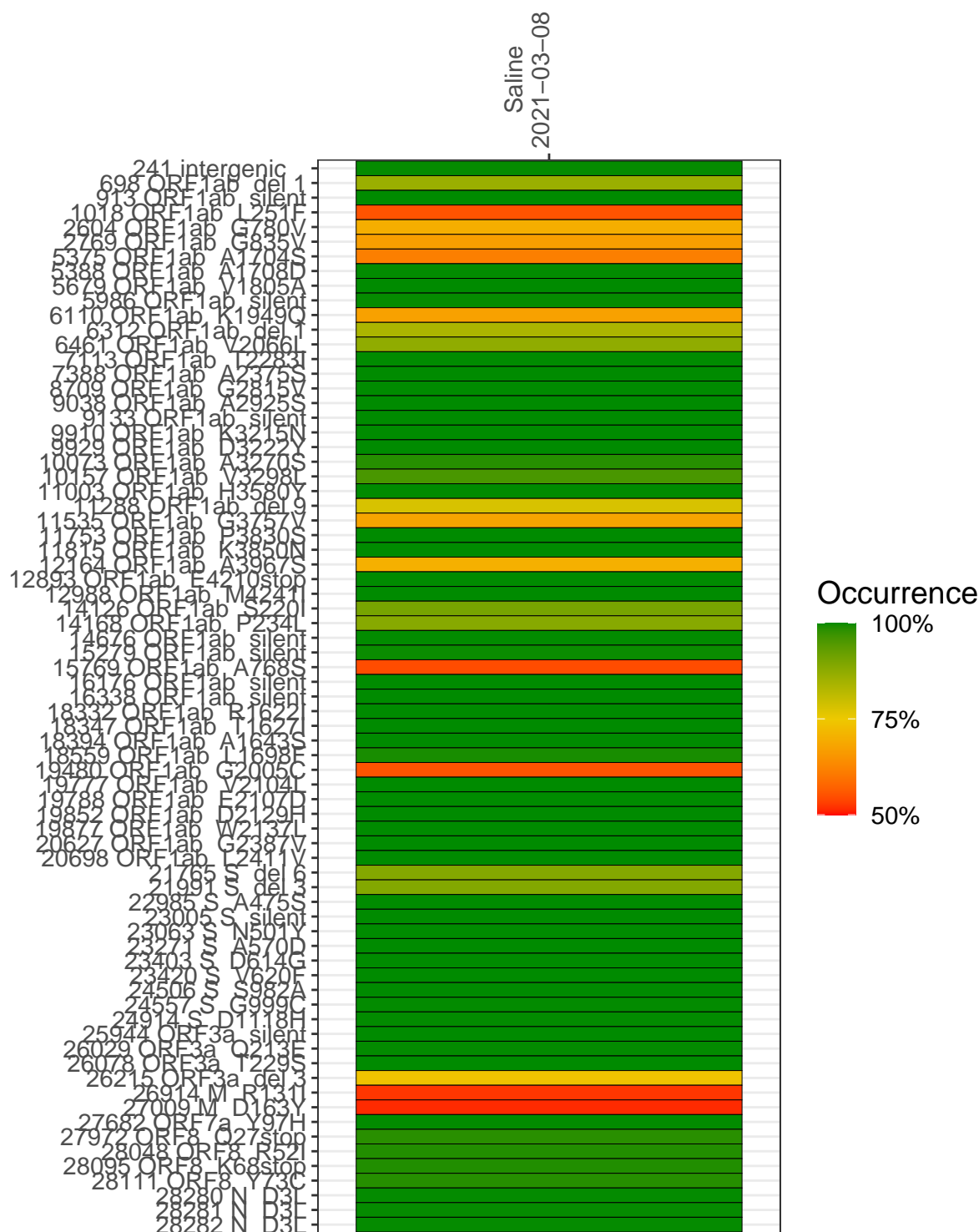
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

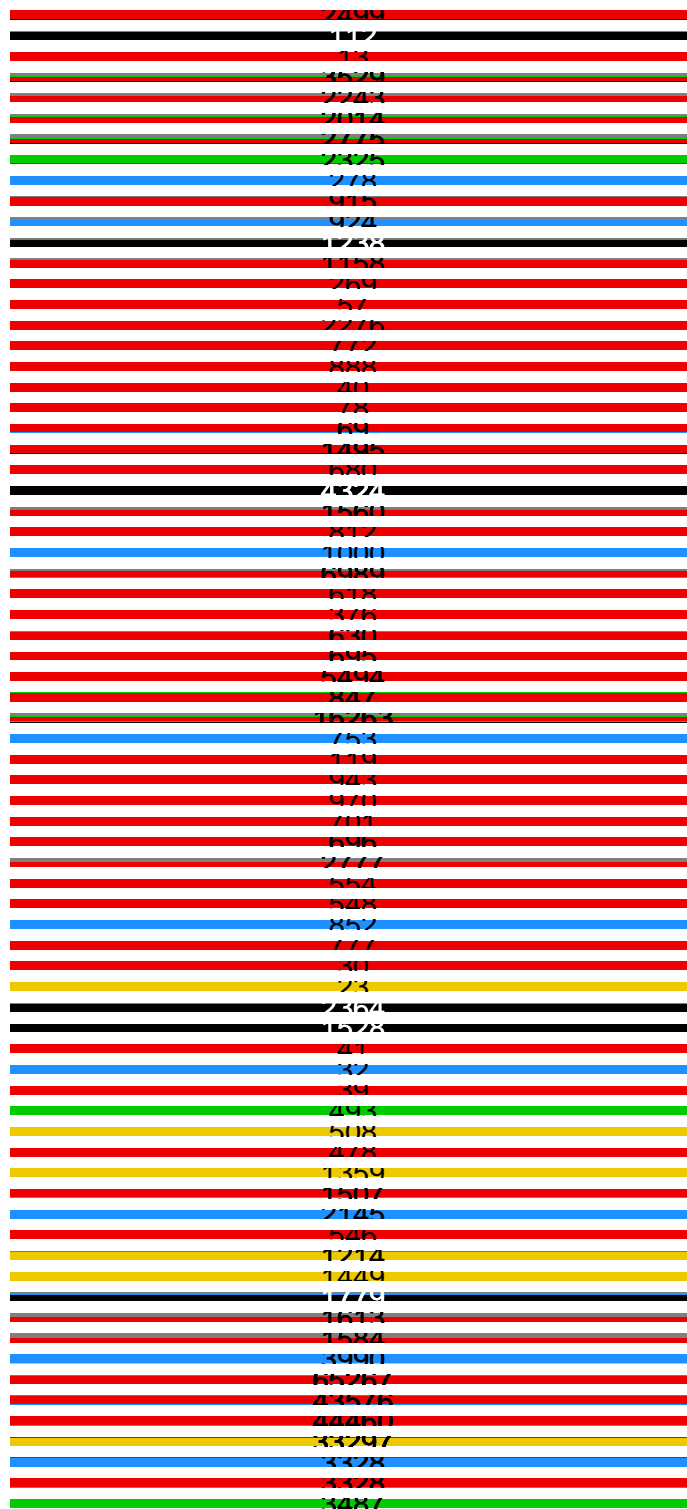
Variants shared across samples

The heat map below shows how variants (reference genome `/home/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
698 CREF12h del 1
913 CREF12h silent
1018 CREF12h T251F
2604 CREF12h G780V
2769 CREF12h G835V
5375 CREF12h A1704S
5388 CREF12h A1708I
5679 CREF12h V1805A
5986 CREF12h silent
6110 CREF12h K1949Q
6312 CREF12h del 1
6461 CREF12h V2066I
7113 CREF12h T2283I
7388 CREF12h A2375S
8709 CREF12h G2815V
9038 CREF12h A2925S
9133 CREF12h silent
9910 CREF12h K3215N
9929 CREF12h T3222Y
10073 CREF12h A3270S
10157 CREF12h V3298I
11003 CREF12h H3580Y
11288 CREF12h del 9
11535 CREF12h G3757V
11753 CREF12h E3830S
11815 CREF12h K3850N
12164 CREF12h A3967S
12893 CREF12h F4210STOP
12988 CREF12h M4241I
14126 CREF12h S220I
14168 CREF12h P234I
14676 CREF12h silent
15279 CREF12h silent
15769 CREF12h A768S
16176 CREF12h silent
16338 CREF12h silent
18332 CREF12h R1622I
18347 CREF12h T1627I
18394 CREF12h A1643S
18559 CREF12h T1698F
19480 CREF12h G2005C
19777 CREF12h V2106I
19788 CREF12h F2107I
19852 CREF12h T2129H
19877 CREF12h W2137I
20627 CREF12h G2387V
20698 CREF12h 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 CREF32 silent
26029 CREF32 C273F
26078 CREF32 T229S
26215 CREF32 del 3
26914 M R131I
27009 M T163Y
27682 CREF32 Y97H
27972 CREF32 C273STOP
28028 CREF32 R52I
28095 CREF32 K68STOP
28111 CREF32 Y73C
28280 N T3I
28281 N T3I
28282 N T3I



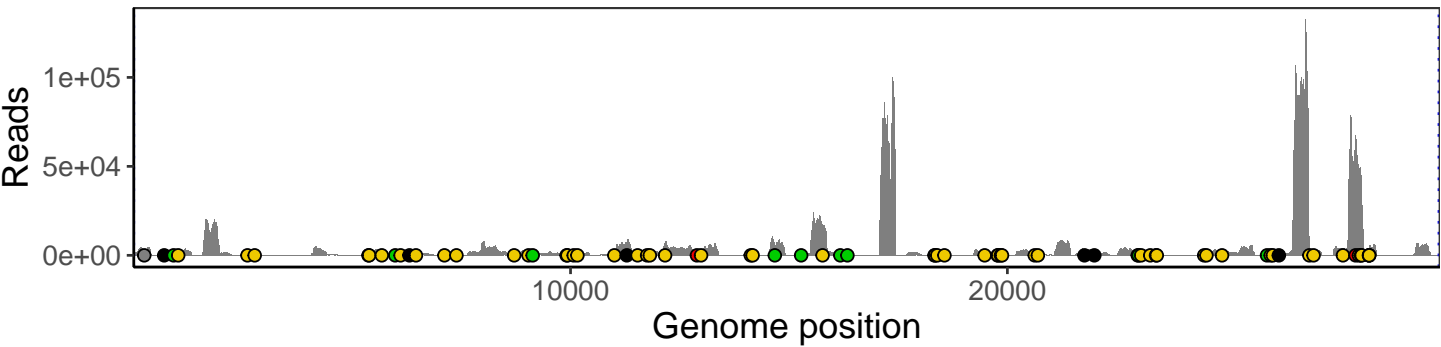
Base change
Expected
A
T
C
G
N
Ins/Del
No data

VSP1032-1

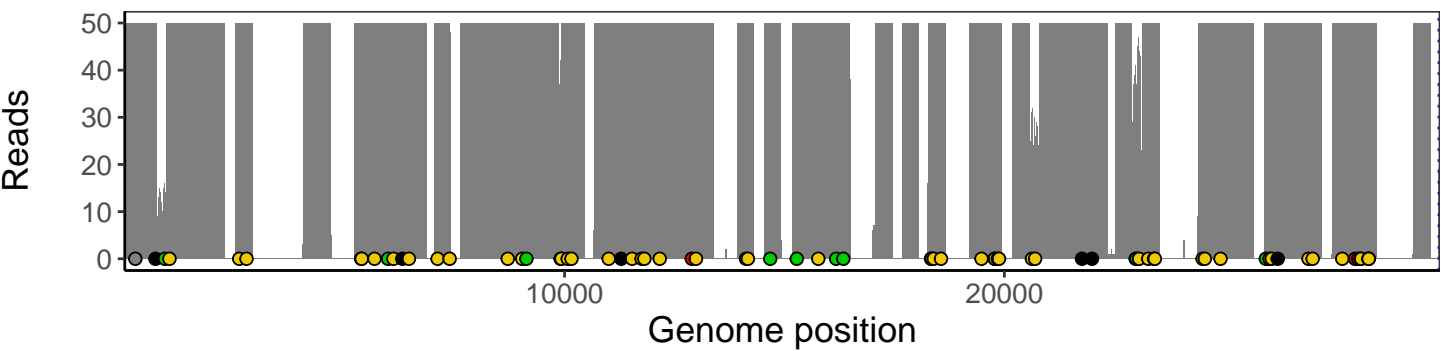
Analyses of individual experiments and composite results

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

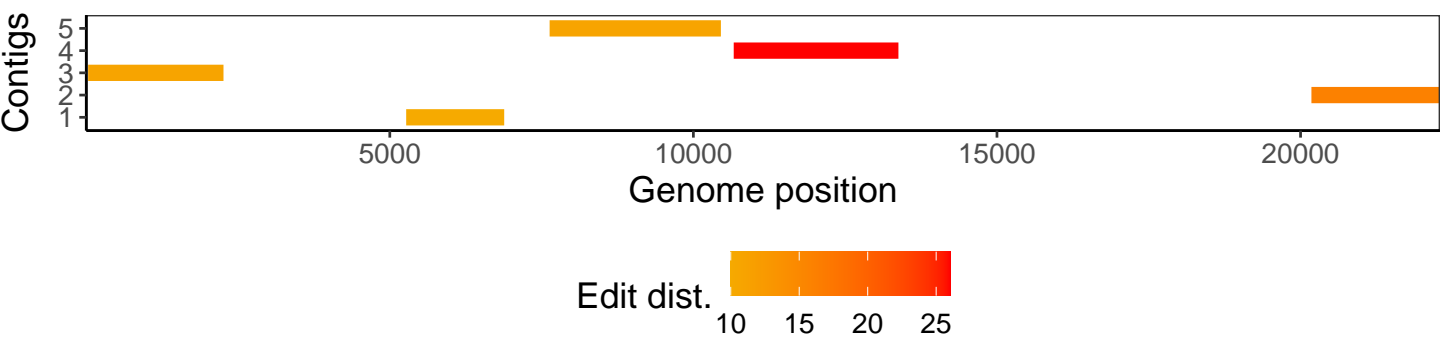
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	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