COVID-19 subject UPHS-0097

2021-06-23

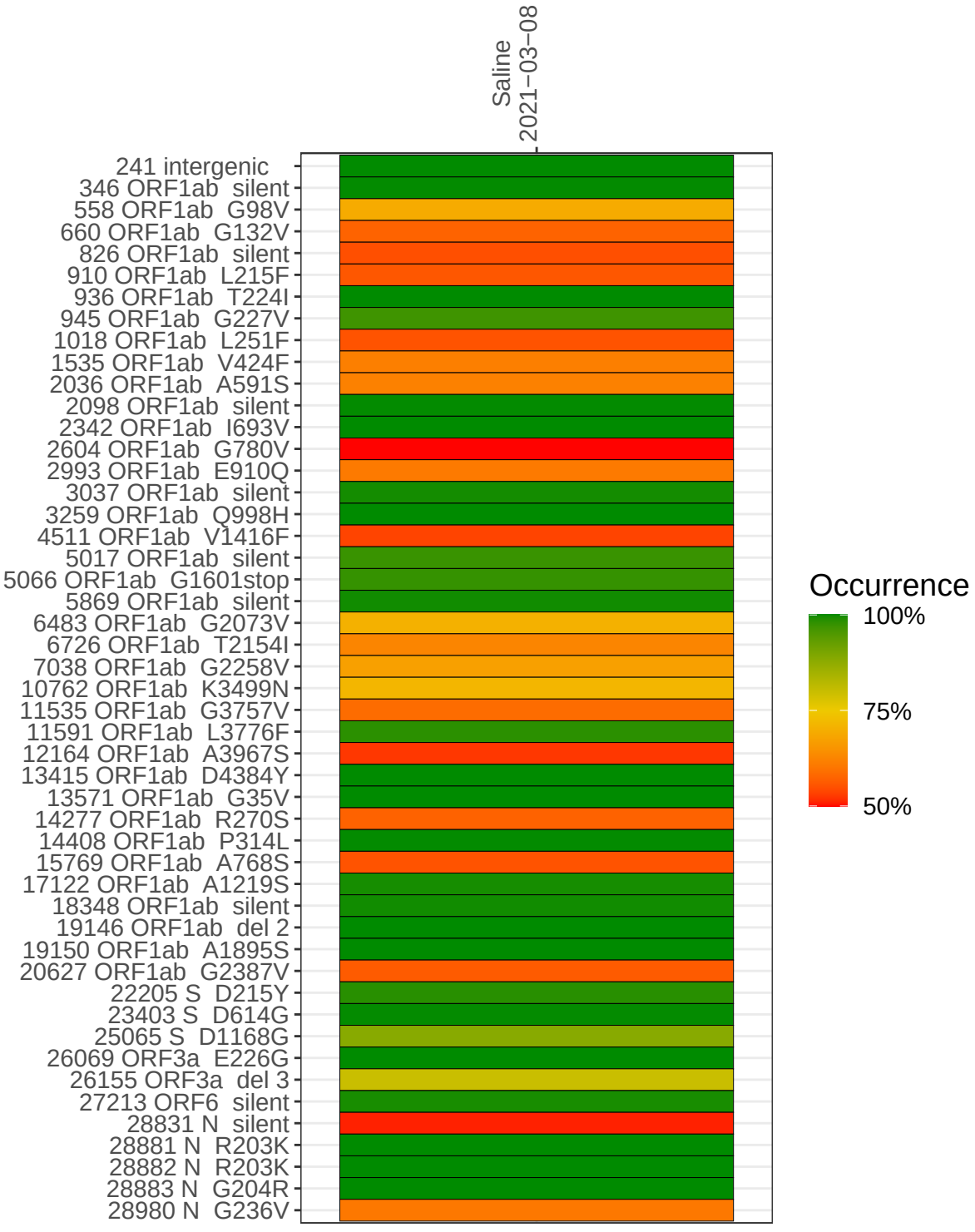
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with > 90% sequence coverage.

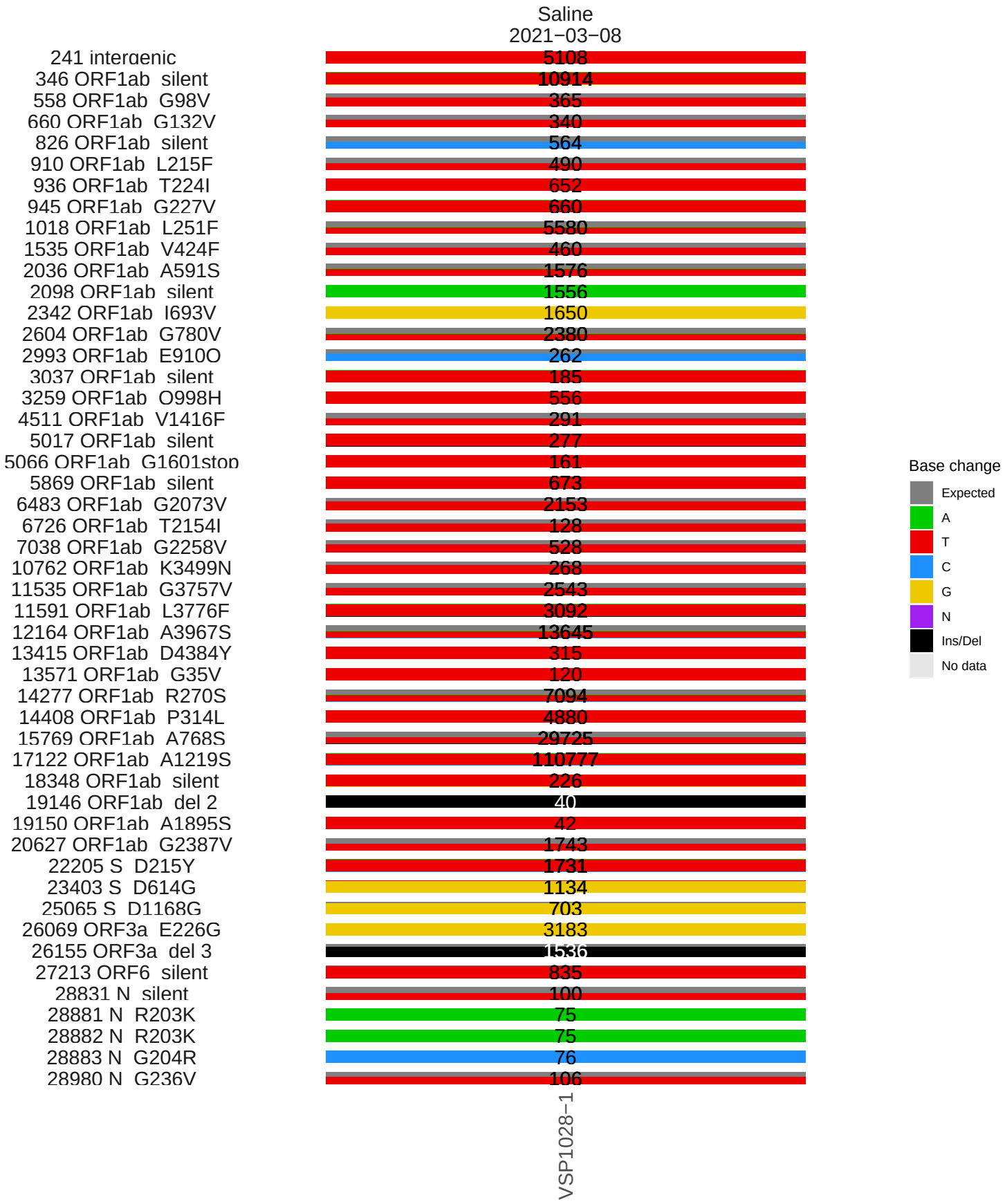
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1028-1	single experiment	NA	Saline	2021-03-08	9.20	NA	92.0%	91.7%

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.

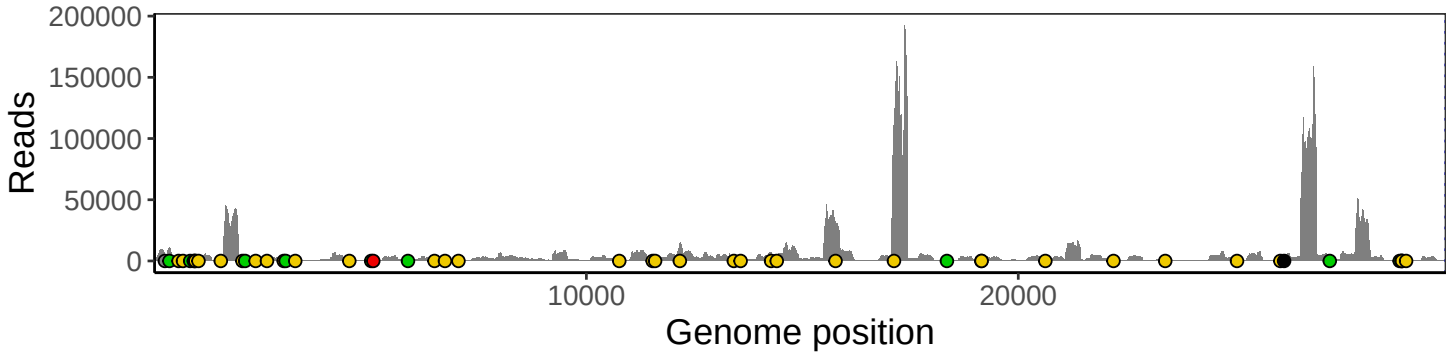



Saline
2021−03−08
241 intergenic
346 ORF1ab silent
558 ORF1ab G98V
660 ORF1ab G132V
826 ORF1ab silent
910 ORF1ab L215F
936 ORF1ab T224I
945 ORF1ab G227V
1018 ORF1ab L251F
1535 ORF1ab V424F
2036 ORF1ab A591S
2098 ORF1ab silent
2342 ORF1ab I693V
2604 ORF1ab G780V
2993 ORF1ab E910O
3037 ORF1ab silent
3259 ORF1ab O998H
4511 ORF1ab V1416F
5017 ORF1ab silent
5066 ORF1ab G1601stop
5869 ORF1ab silent
6483 ORF1ab G2073V
6726 ORF1ab T2154I
7038 ORF1ab G2258V
10762 ORF1ab K3499N
11535 ORF1ab G3757V
11591 ORF1ab L3776F
12164 ORF1ab A3967S
13415 ORF1ab D4384Y
13571 ORF1ab G35V
14277 ORF1ab R270S
14408 ORF1ab P314L
15769 ORF1ab A768S
17122 ORF1ab A1219S
18348 ORF1ab silent
19146 ORF1ab del 2
19150 ORF1ab A1895S
20627 ORF1ab G2387V
22205 S D215Y
23403 S D614G
25065 S D1168G
26069 ORF3a E226G
26155 ORF3a del 3
27213 ORF6 silent
28831 N silent
28881 N R203K
28882 N R203K
28883 N G204R
28980 N G236V
5108
10914
365
340
564
490
652
660
5580
460
1576
1556
1650
2380
262
185
556
291
277
161
673
2153
128
528
268
2543
3092
13645
315
120
7094
4880
29725
110777
226
40
42
1743
1731
1134
703
3183
1536
835
100
75
75
76
106
VSP1028−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

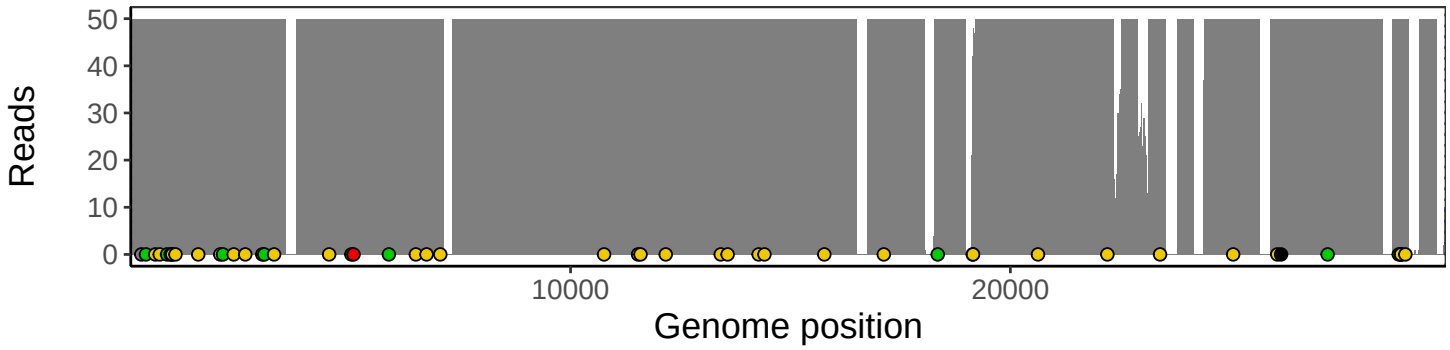
Analyses of individual experiments and composite results

VSP1028-1 | 2021-03-08 | Saline | UPHS-0097 | genomes | single experiment

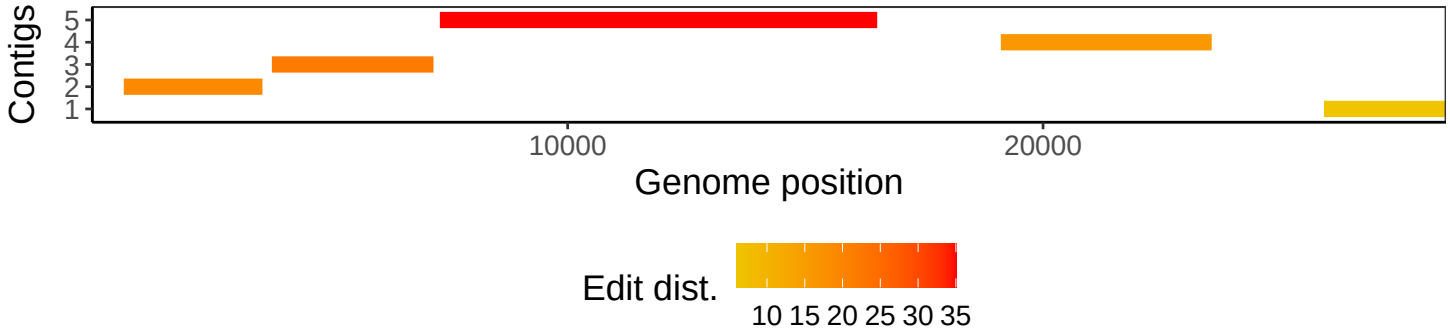
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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 htslib 1.10
bcftools	1.10.2-34-g1a12af0-dirty Using htslib 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