COVID-19 subject UPHS-0647

2021-06-03

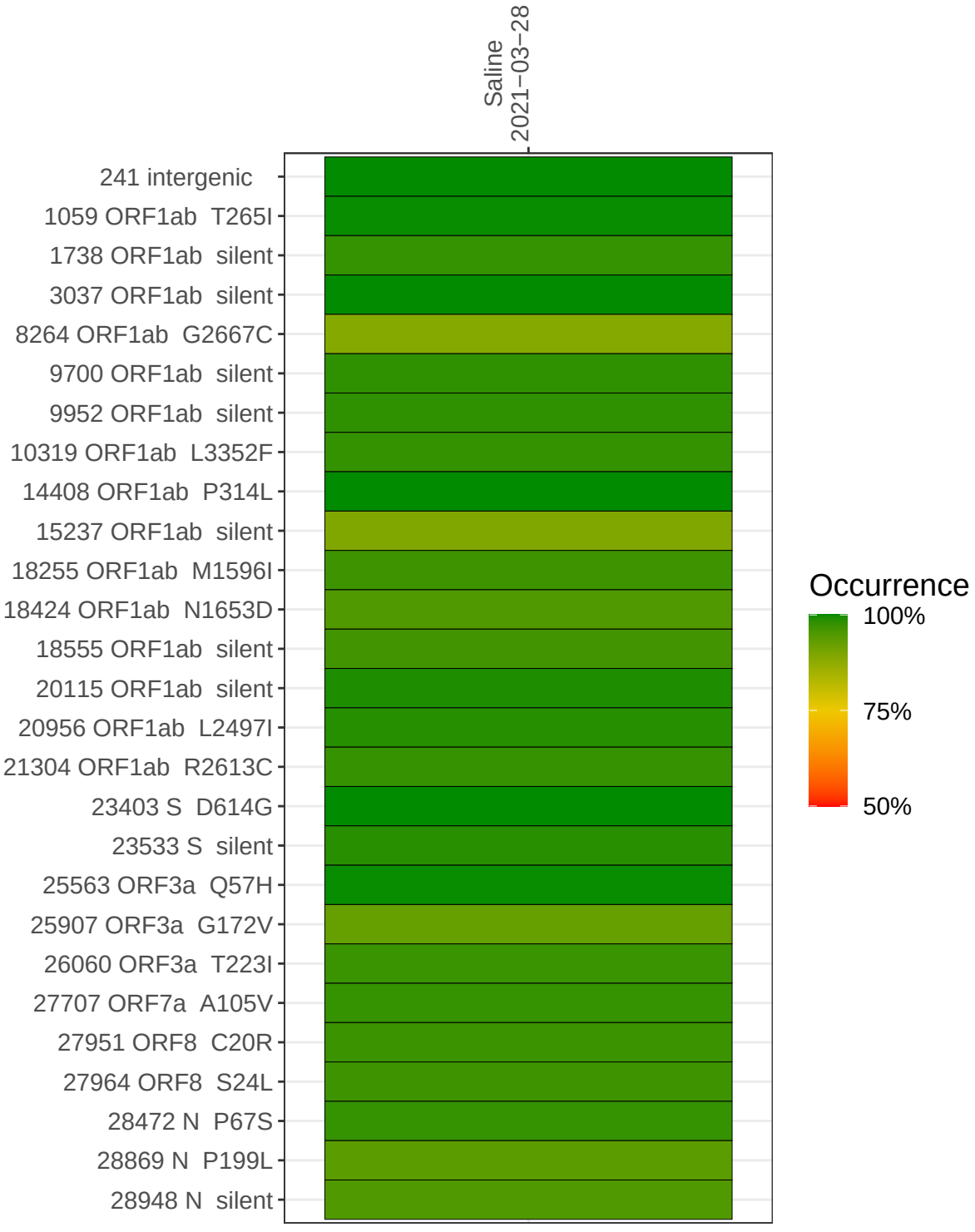
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

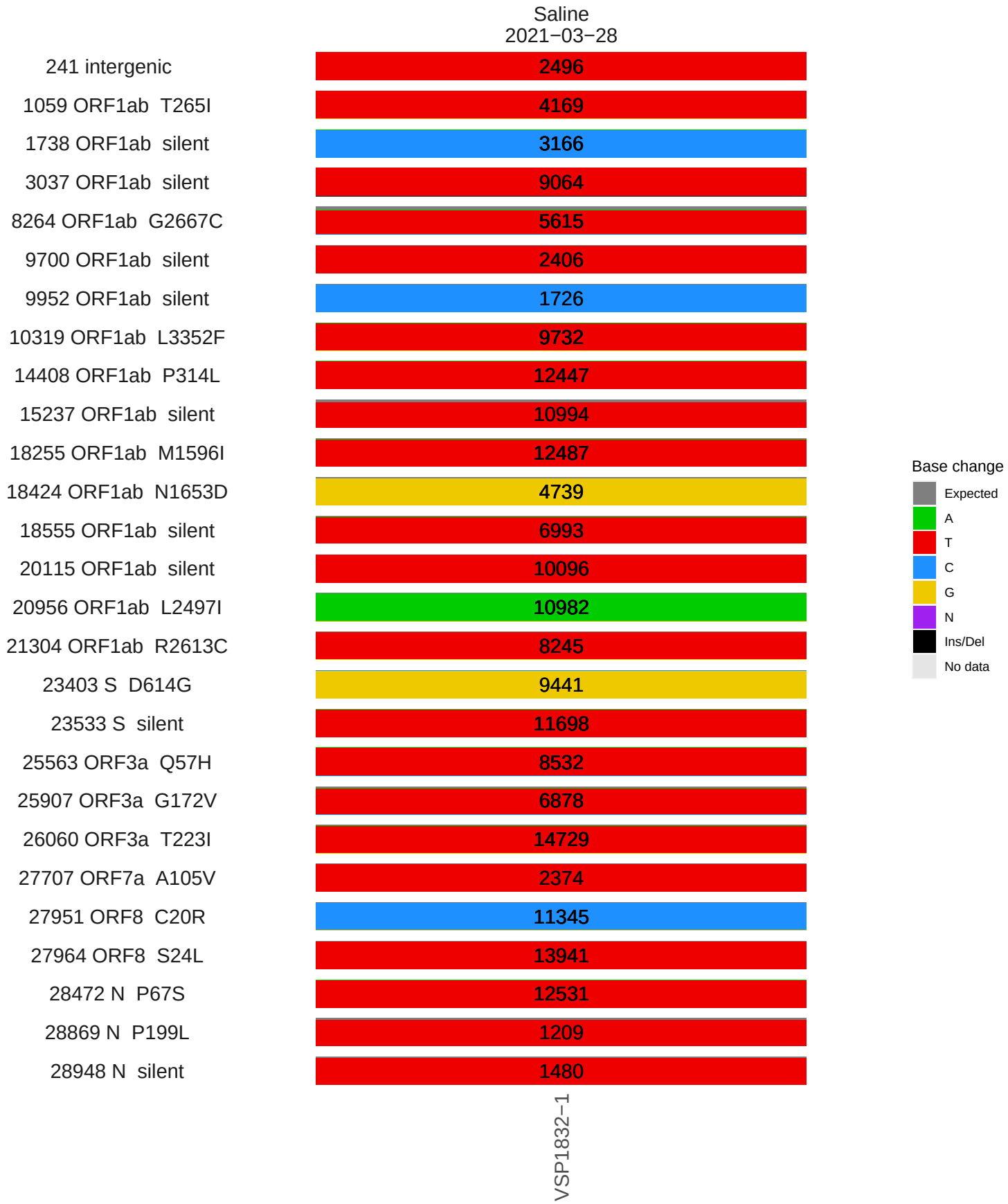
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1832-1	single experiment	NA	Saline	2021-03-28	22.40	B.1.2	100.0%	99.9%

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.

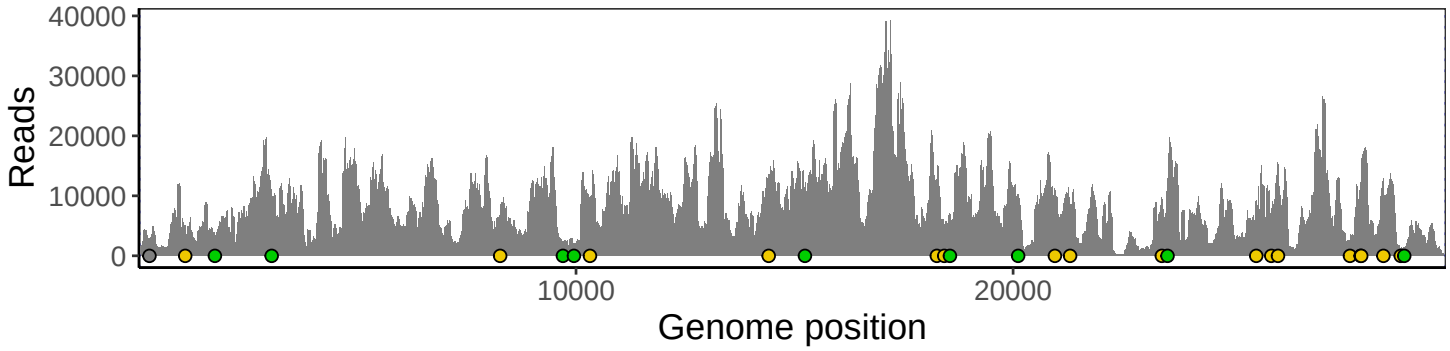



Saline
2021−03−28
241 intergenic
2496
1059 ORF1ab T265I
4169
1738 ORF1ab silent
3166
3037 ORF1ab silent
9064
8264 ORF1ab G2667C
5615
9700 ORF1ab silent
2406
9952 ORF1ab silent
1726
10319 ORF1ab L3352F
9732
14408 ORF1ab P314L
12447
15237 ORF1ab silent
10994
18255 ORF1ab M1596I
12487
18424 ORF1ab N1653D
4739
18555 ORF1ab silent
6993
20115 ORF1ab silent
10096
20956 ORF1ab L2497I
10982
21304 ORF1ab R2613C
8245
23403 S D614G
9441
23533 S silent
11698
25563 ORF3a Q57H
8532
25907 ORF3a G172V
6878
26060 ORF3a T223I
14729
27707 ORF7a A105V
2374
27951 ORF8 C20R
11345
27964 ORF8 S24L
13941
28472 N P67S
12531
28869 N P199L
1209
28948 N silent
1480
VSP1832−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

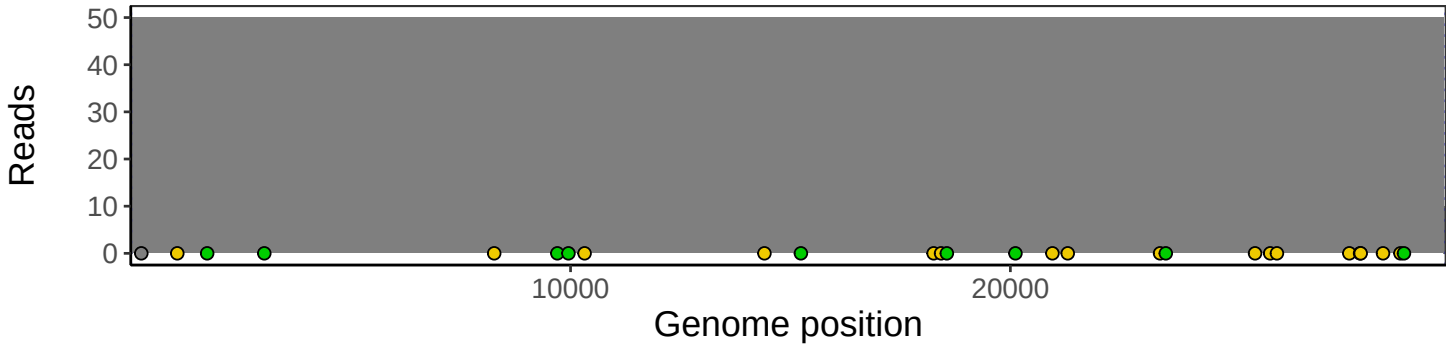
Analyses of individual experiments and composite results

VSP1832-1 | 2021-03-28 | Saline | UPHS-0647 | genomes | single experiment

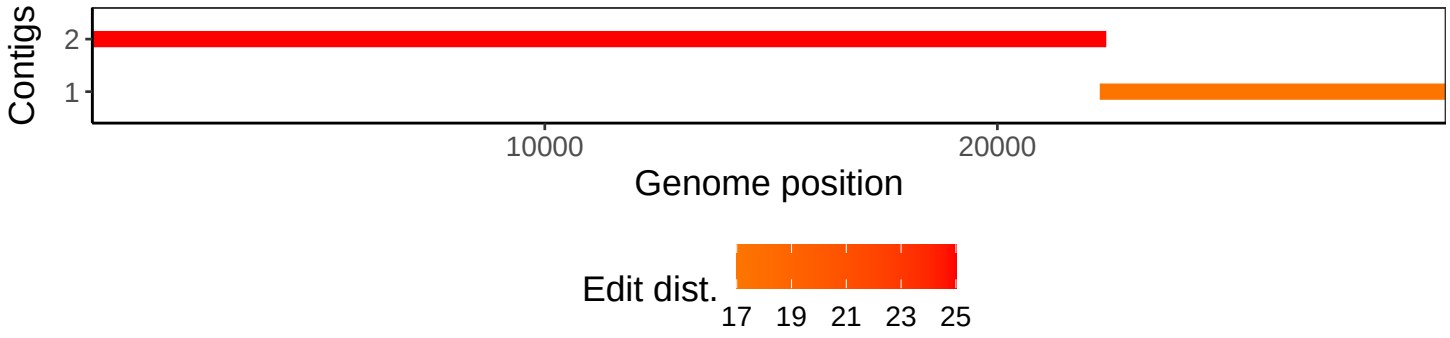
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	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