COVID-19 subject UPHS-0030

2021-03-25

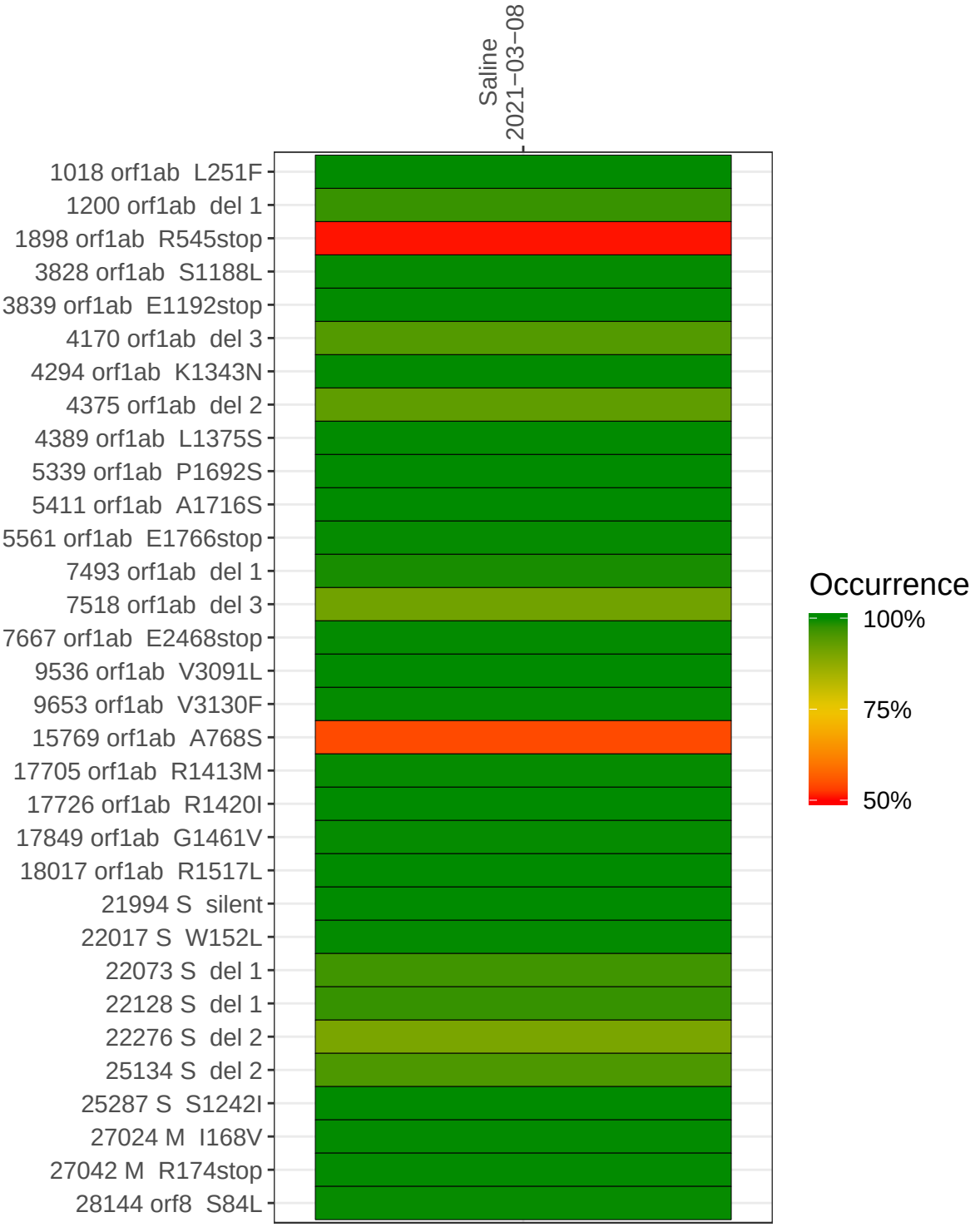
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

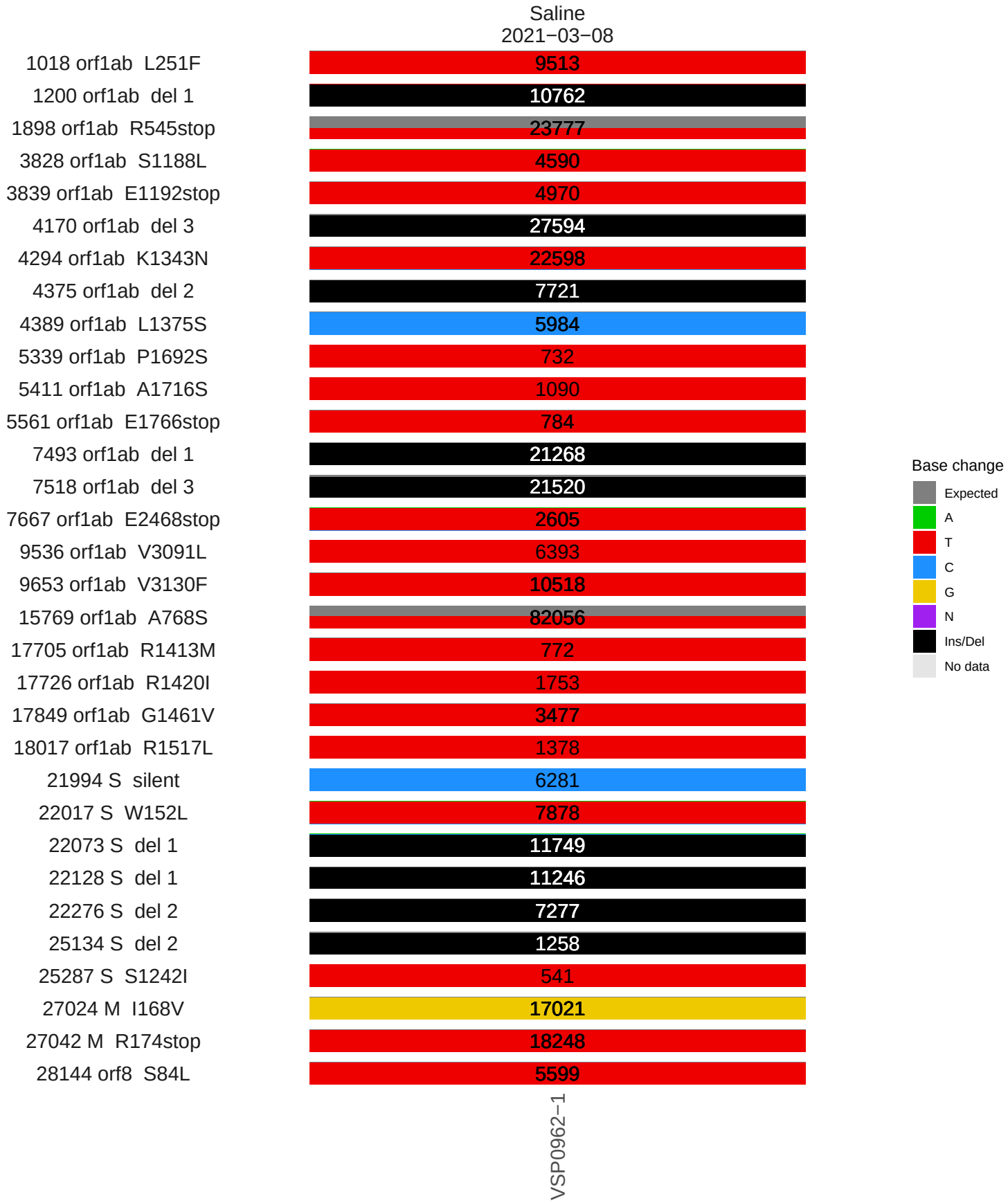
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0962-1	single experiment	NA	Saline	2021-03-08	0.72	NA	21.4%	19.6%

Variants shared across samples

The heat map below shows how variants (reference genome USA-WA1-2020) 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.

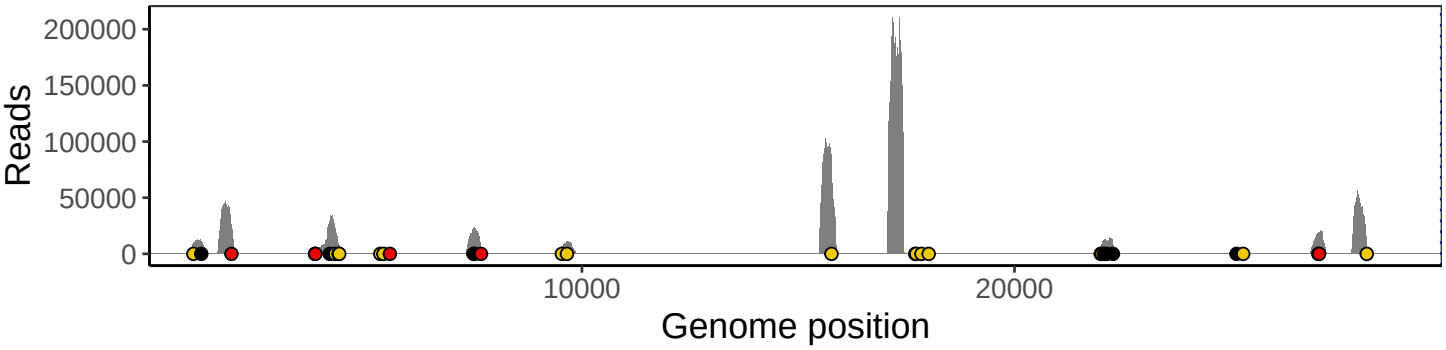



Saline
2021−03−08
1018 orf1ab L251F
9513
1200 orf1ab del 1
10762
1898 orf1ab R545stop
23777
3828 orf1ab S1188L
4590
3839 orf1ab E1192stop
4970
4170 orf1ab del 3
27594
4294 orf1ab K1343N
22598
4375 orf1ab del 2
7721
4389 orf1ab L1375S
5984
5339 orf1ab P1692S
732
5411 orf1ab A1716S
1090
5561 orf1ab E1766stop
784
7493 orf1ab del 1
21268
7518 orf1ab del 3
21520
7667 orf1ab E2468stop
2605
9536 orf1ab V3091L
6393
9653 orf1ab V3130F
10518
15769 orf1ab A768S
82056
17705 orf1ab R1413M
772
17726 orf1ab R1420I
1753
17849 orf1ab G1461V
3477
18017 orf1ab R1517L
1378
21994 S silent
6281
22017 S W152L
7878
22073 S del 1
11749
22128 S del 1
11246
22276 S del 2
7277
25134 S del 2
1258
25287 S S1242I
541
27024 M I168V
17021
27042 M R174stop
18248
28144 orf8 S84L
5599
VSP0962−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

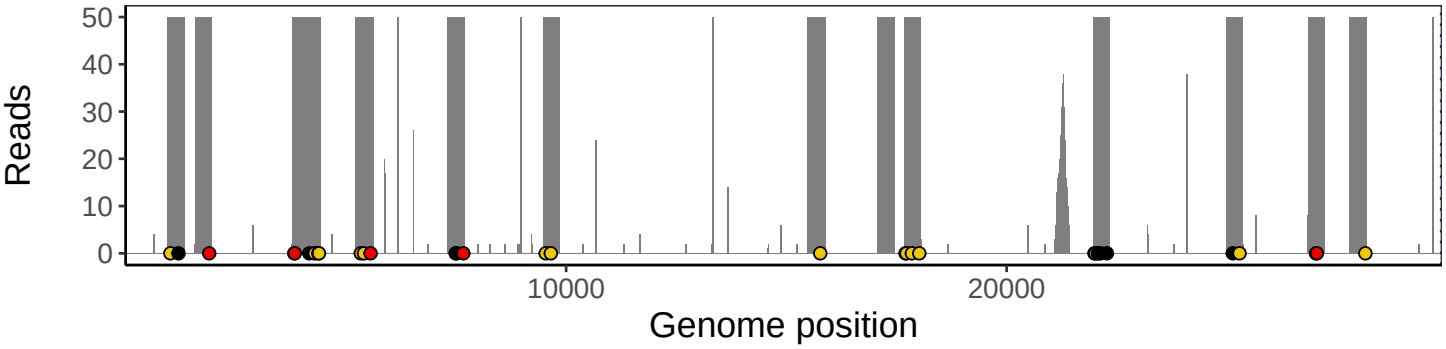
Analyses of individual experiments and composite results

VSP0962-1 | 2021-03-08 | Saline | UPHS-0030 | genomes | single experiment

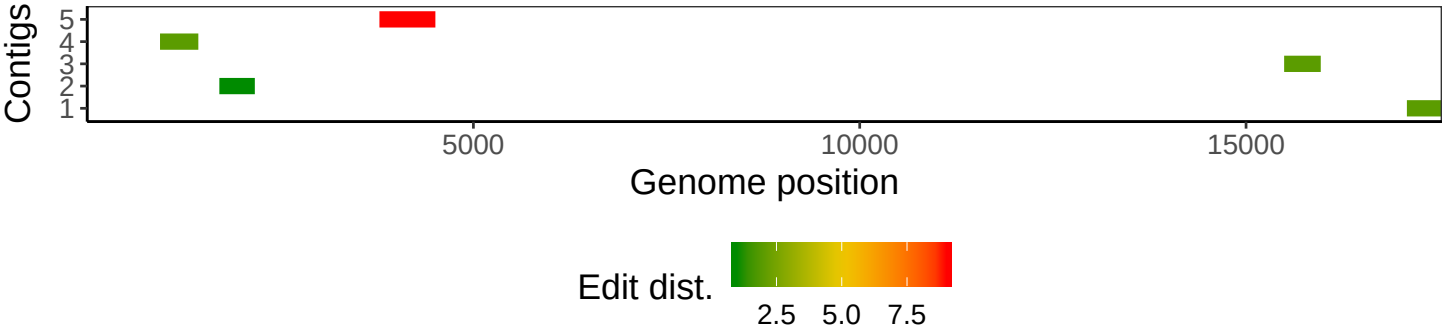
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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 htslib 1.10
bcftools	1.10.2-34-g1a12af0-dirty Using htslib 1.10.2-57-gf58a6f3
pangolin	2.3.3
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	3.0.0
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