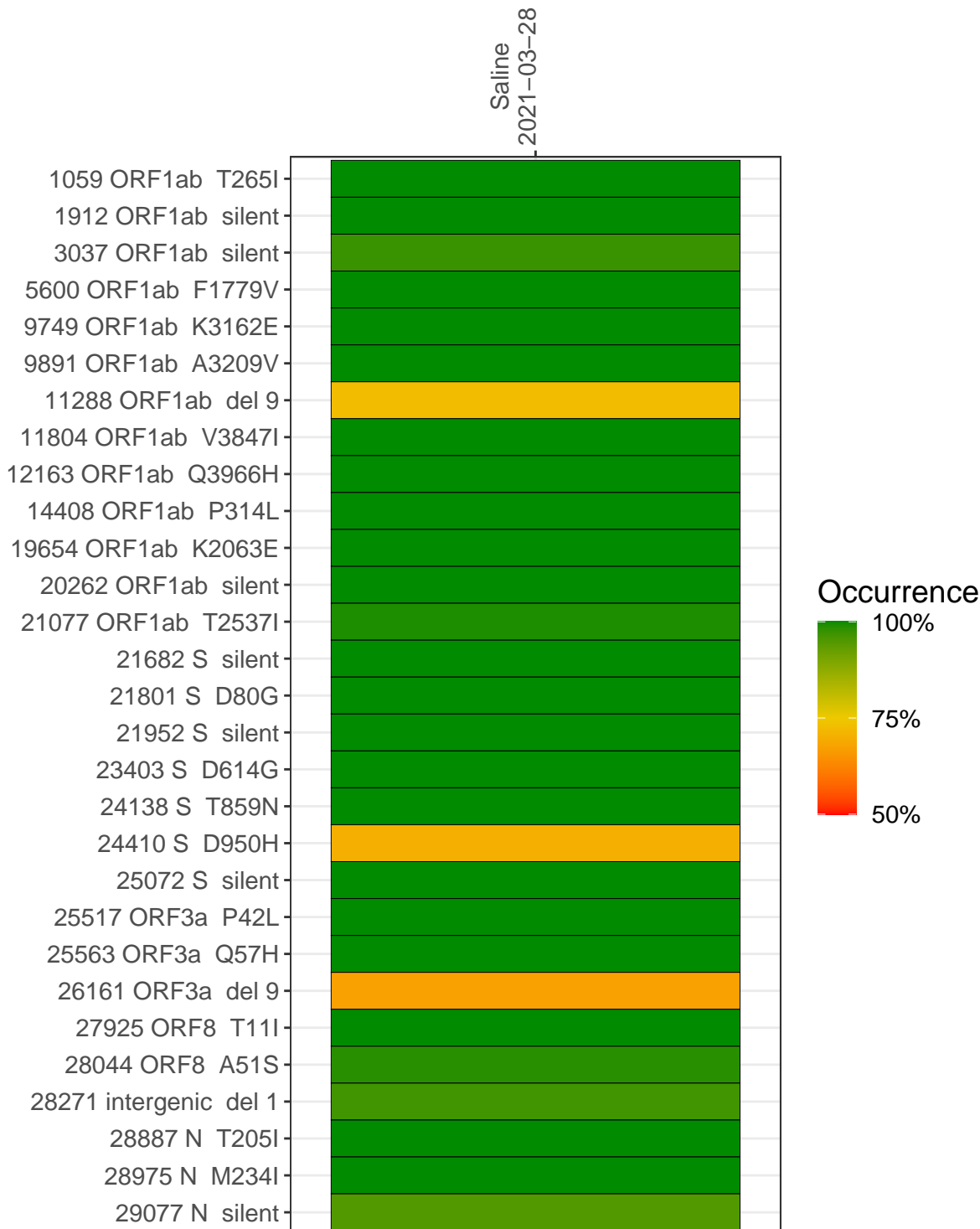# COVID-19 subject UPHS-0622

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.
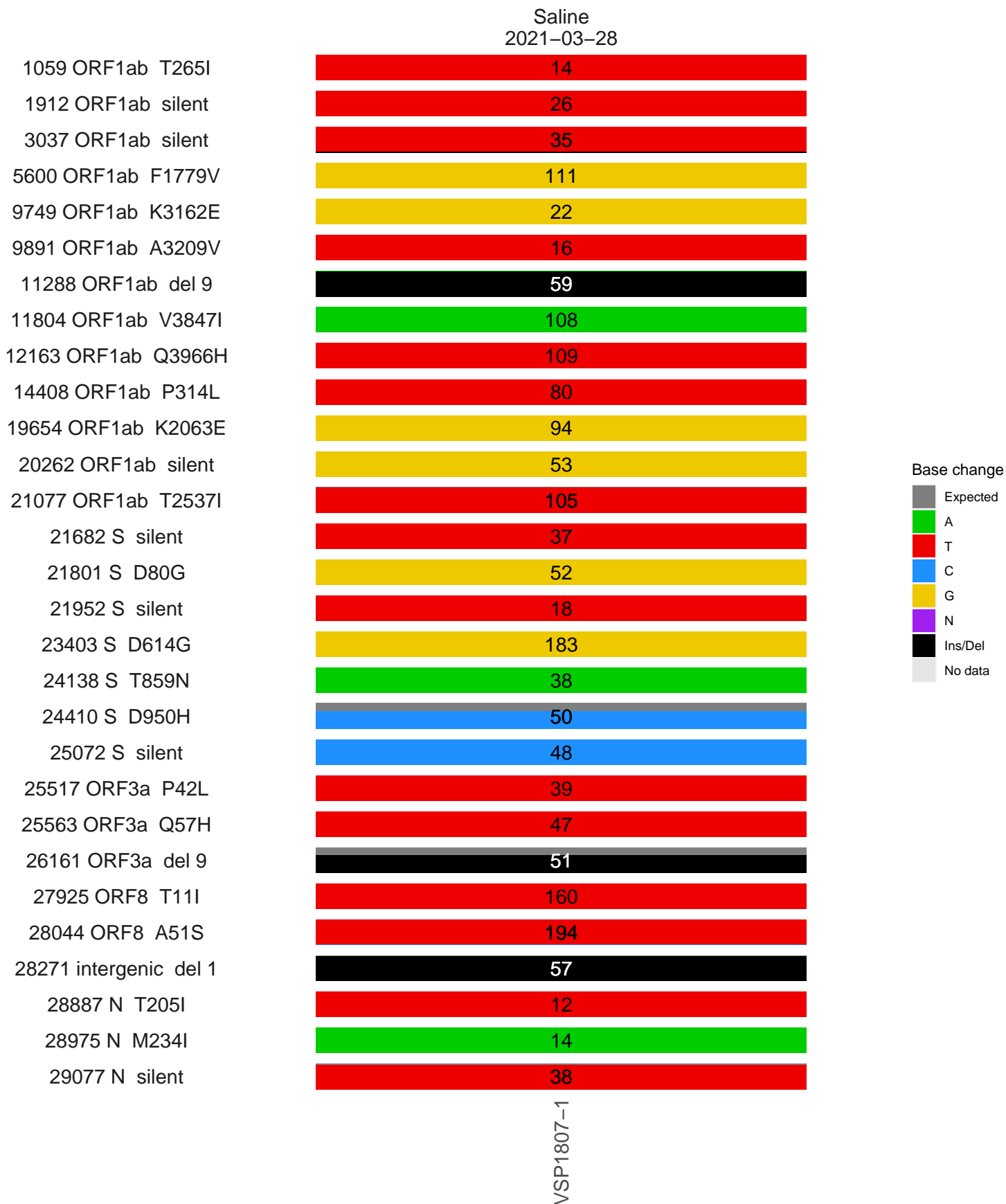
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1807-1	single experiment	NA	Saline	2021-03-28	17.13	B.1	99.5%	97.4%

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

Position	VSP1807-1
1059 ORF1ab T265I	14
1912 ORF1ab silent	26
3037 ORF1ab silent	35
5600 ORF1ab F1779V	111
9749 ORF1ab K3162E	22
9891 ORF1ab A3209V	16
11288 ORF1ab del 9	59
11804 ORF1ab V3847I	108
12163 ORF1ab Q3966H	109
14408 ORF1ab P314L	80
19654 ORF1ab K2063E	94
20262 ORF1ab silent	53
21077 ORF1ab T2537I	105
21682 S silent	37
21801 S D80G	52
21952 S silent	18
23403 S D614G	183
24138 S T859N	38
24410 S D950H	50
25072 S silent	48
25517 ORF3a P42L	39
25563 ORF3a Q57H	47
26161 ORF3a del 9	51
27925 ORF8 T11I	160
28044 ORF8 A51S	194
28271 intergenic del 1	57
28887 N T205I	12
28975 N M234I	14
29077 N silent	38

Base change

- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

Analyses of individual experiments and composite results

**VSP1807-1 | 2021-03-28 | Saline | UPHS-0622 | 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.
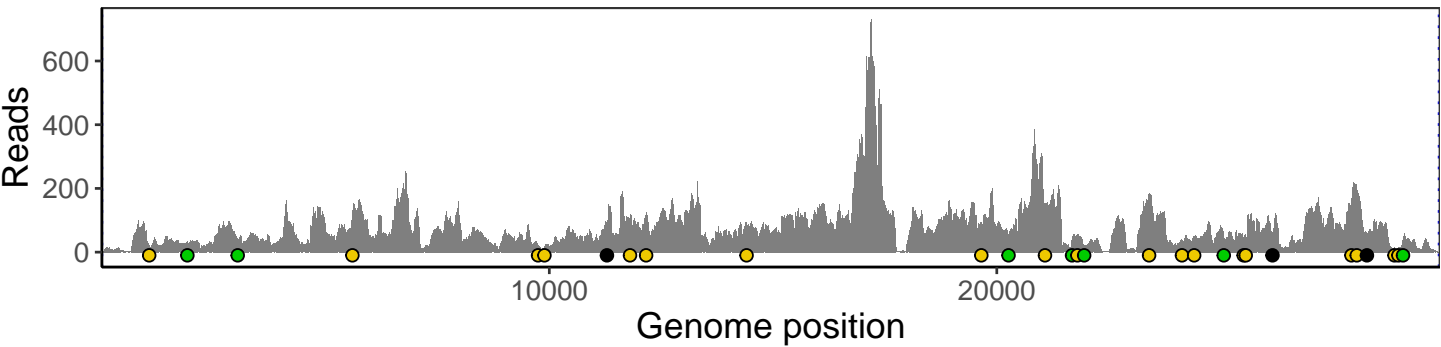
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
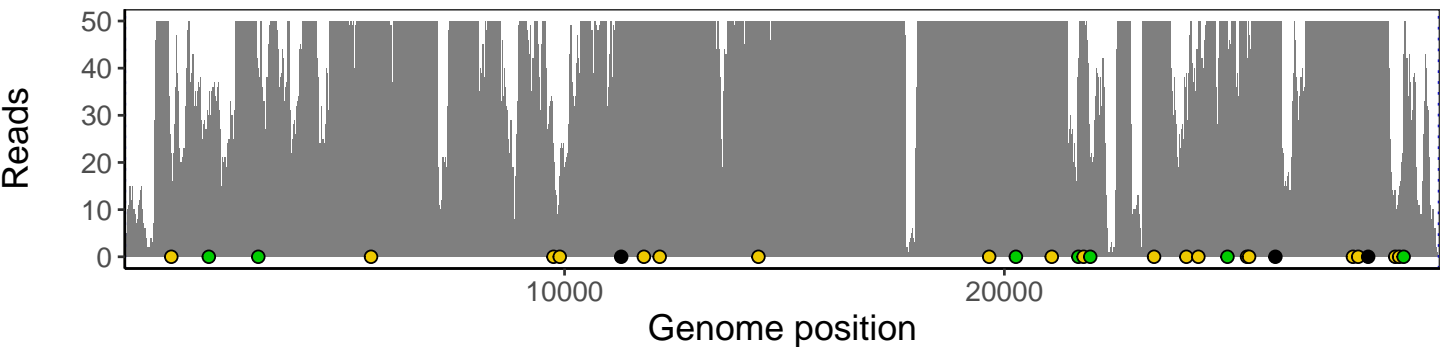
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

4
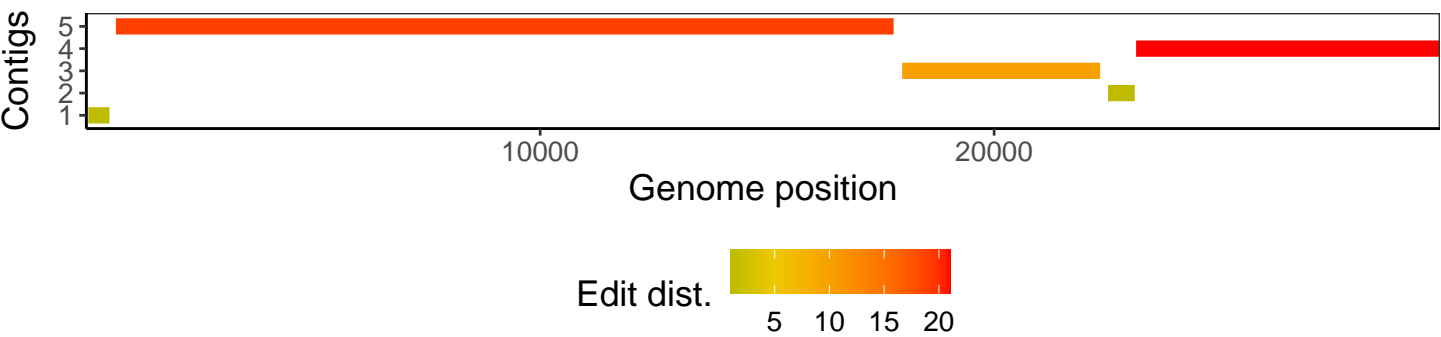
Software environment

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