# COVID-19 subject UPHS-1644

*2021-06-03*

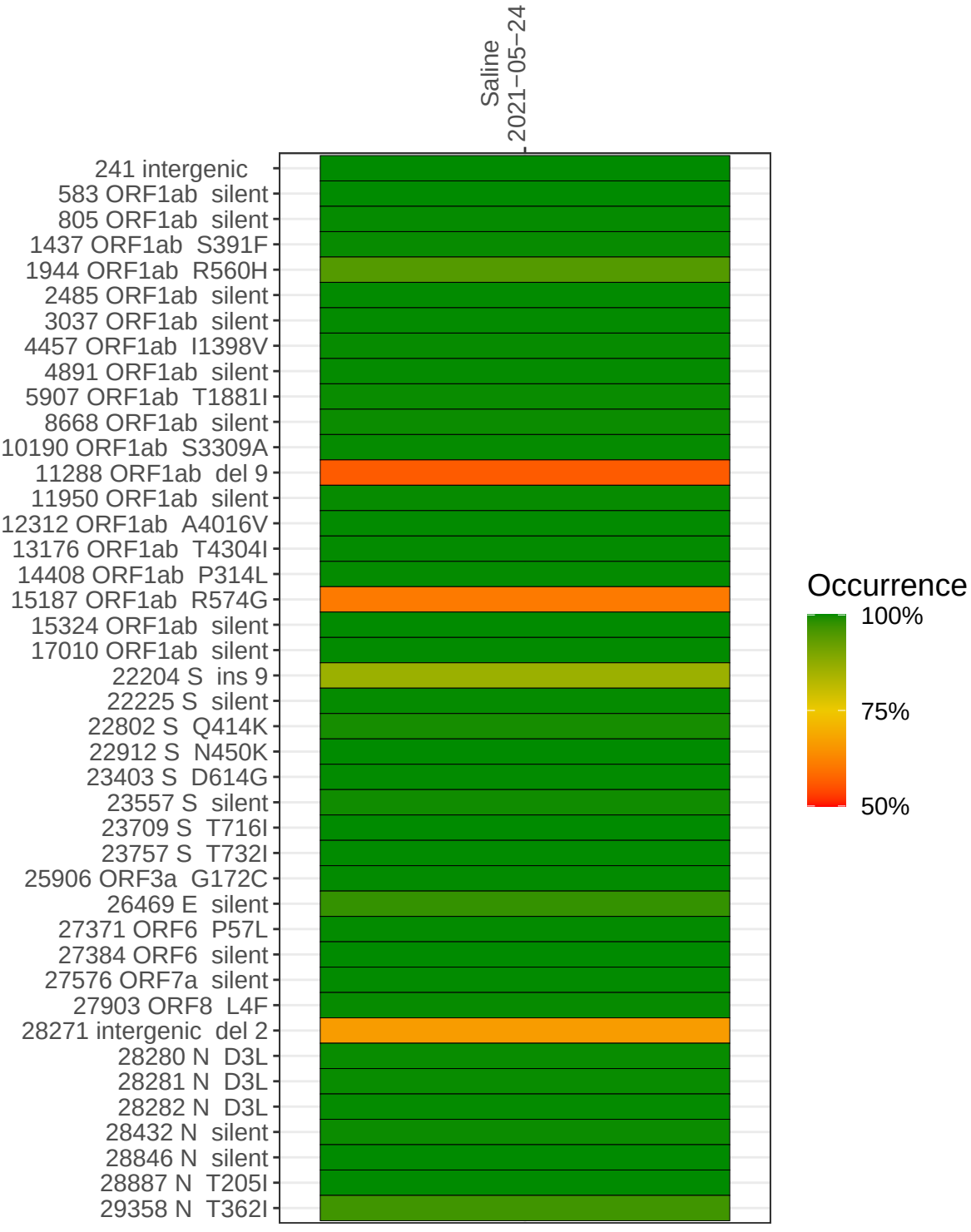
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

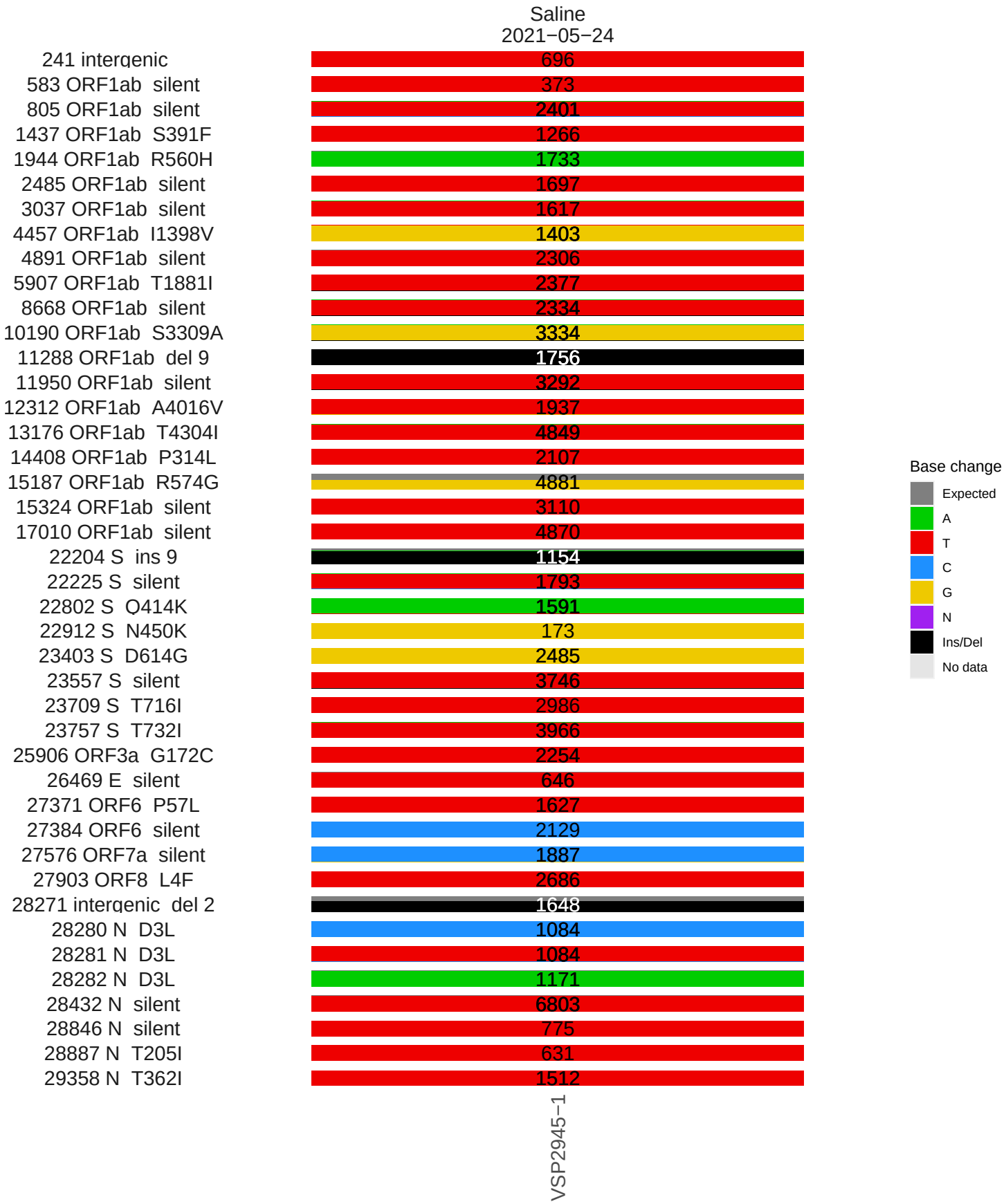
Table 1. Sample summary.

| Experiment | Type | Genomes | Sample type | Sample date | Largest contig (KD) | Lineage | Reference read coverage | Reference read coverage (>= 5 reads) |
|---|---|---|---|---|---|---|---|---|
| VSP2945-1 | single experiment | NA | Saline | 2021-05-24 | 29.80 | B.1.214.2 | 99.8% | 99.8% |

**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.

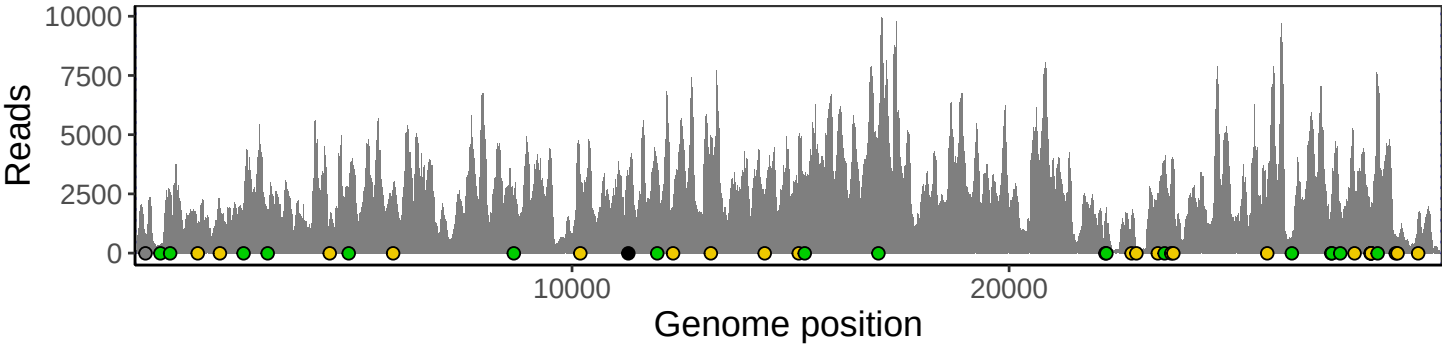



Saline
2021−05−24
241 intergenic
583 ORF1ab silent
805 ORF1ab silent
1437 ORF1ab S391F
1944 ORF1ab R560H
2485 ORF1ab silent
3037 ORF1ab silent
4457 ORF1ab I1398V
4891 ORF1ab silent
5907 ORF1ab T1881I
8668 ORF1ab silent
10190 ORF1ab S3309A
11288 ORF1ab del 9
11950 ORF1ab silent
12312 ORF1ab A4016V
13176 ORF1ab T4304I
14408 ORF1ab P314L
15187 ORF1ab R574G
15324 ORF1ab silent
17010 ORF1ab silent
22204 S ins 9
22225 S silent
22802 S Q414K
22912 S N450K
23403 S D614G
23557 S silent
23709 S T716I
23757 S T732I
25906 ORF3a G172C
26469 E silent
27371 ORF6 P57L
27384 ORF6 silent
27576 ORF7a silent
27903 ORF8 L4F
28271 intergenic del 2
28280 N D3L
28281 N D3L
28282 N D3L
28432 N silent
28846 N silent
28887 N T205I
29358 N T362I
696
373
2401
1266
1733
1697
1617
1403
2306
2377
2334
3334
1756
3292
1937
4849
2107
4881
3110
4870
1154
1793
1591
173
2485
3746
2986
3966
2254
646
1627
2129
1887
2686
1648
1084
1084
1171
6803
775
631
1512
VSP2945−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

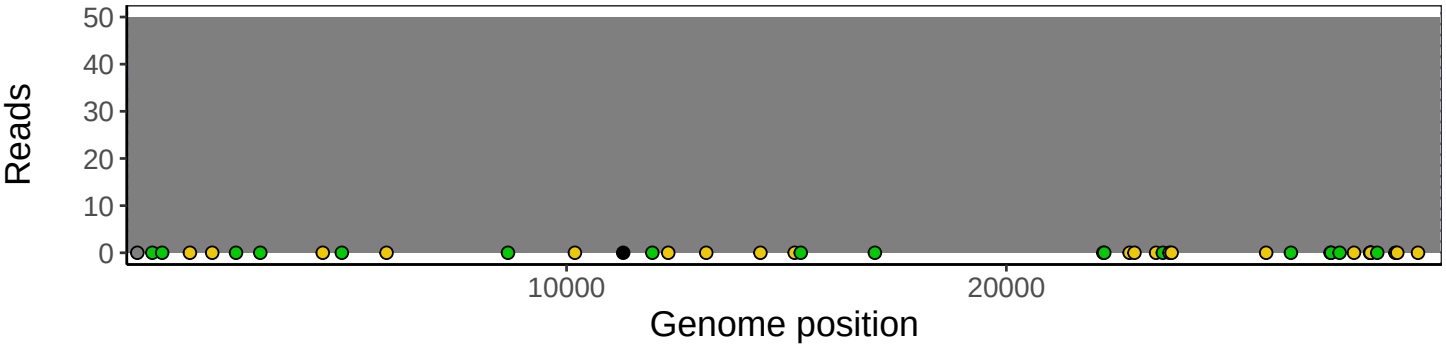
## Analyses of individual experiments and composite results

### VSP2945-1 | 2021-05-24 | Saline | UPHS-1644 | genomes | single experiment

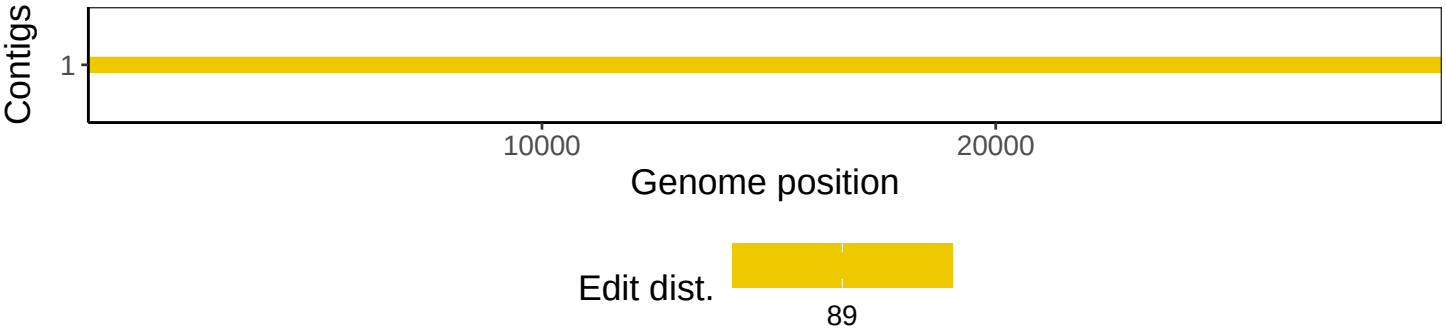
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.




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 |