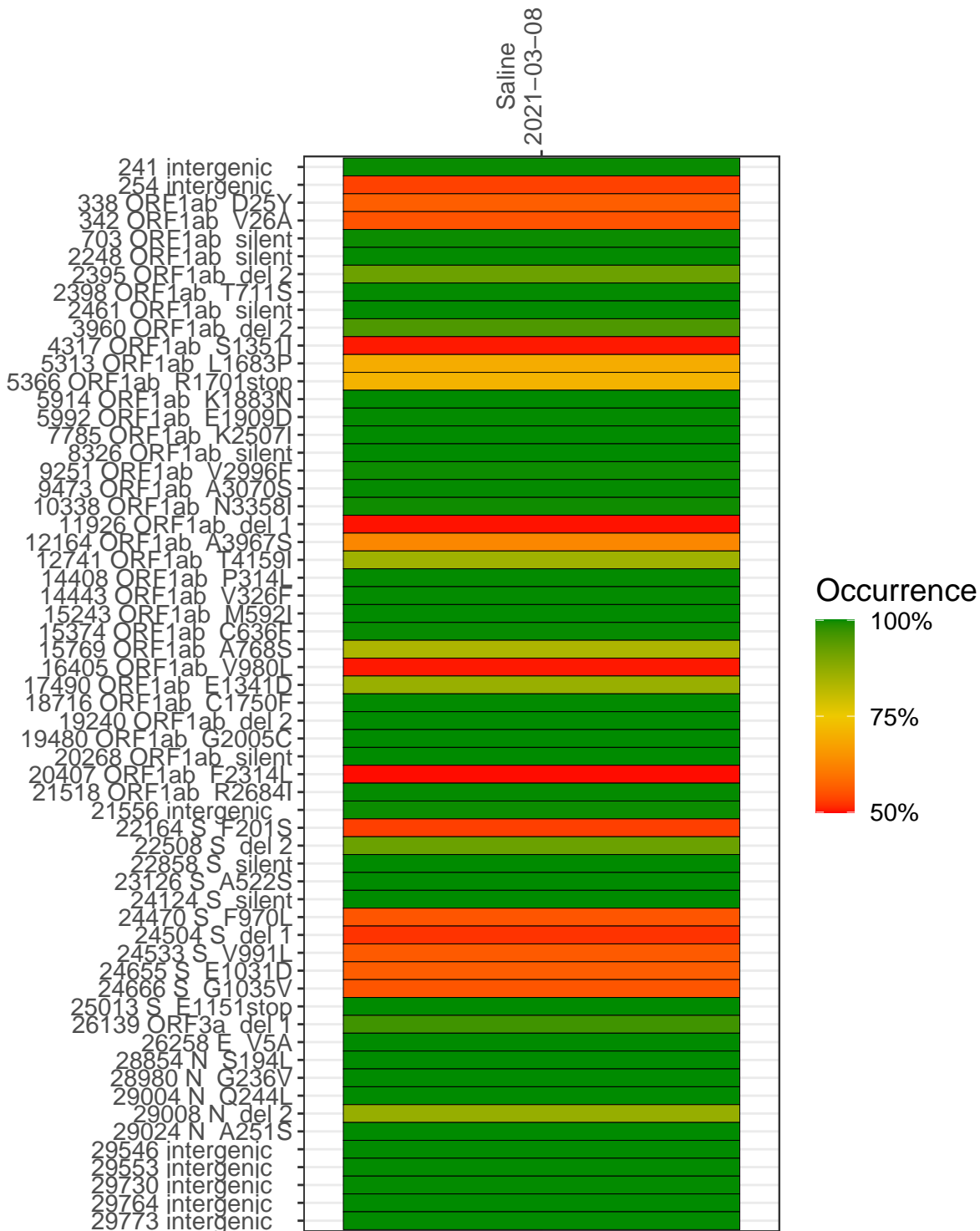# COVID-19 subject UPHS-0037

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.
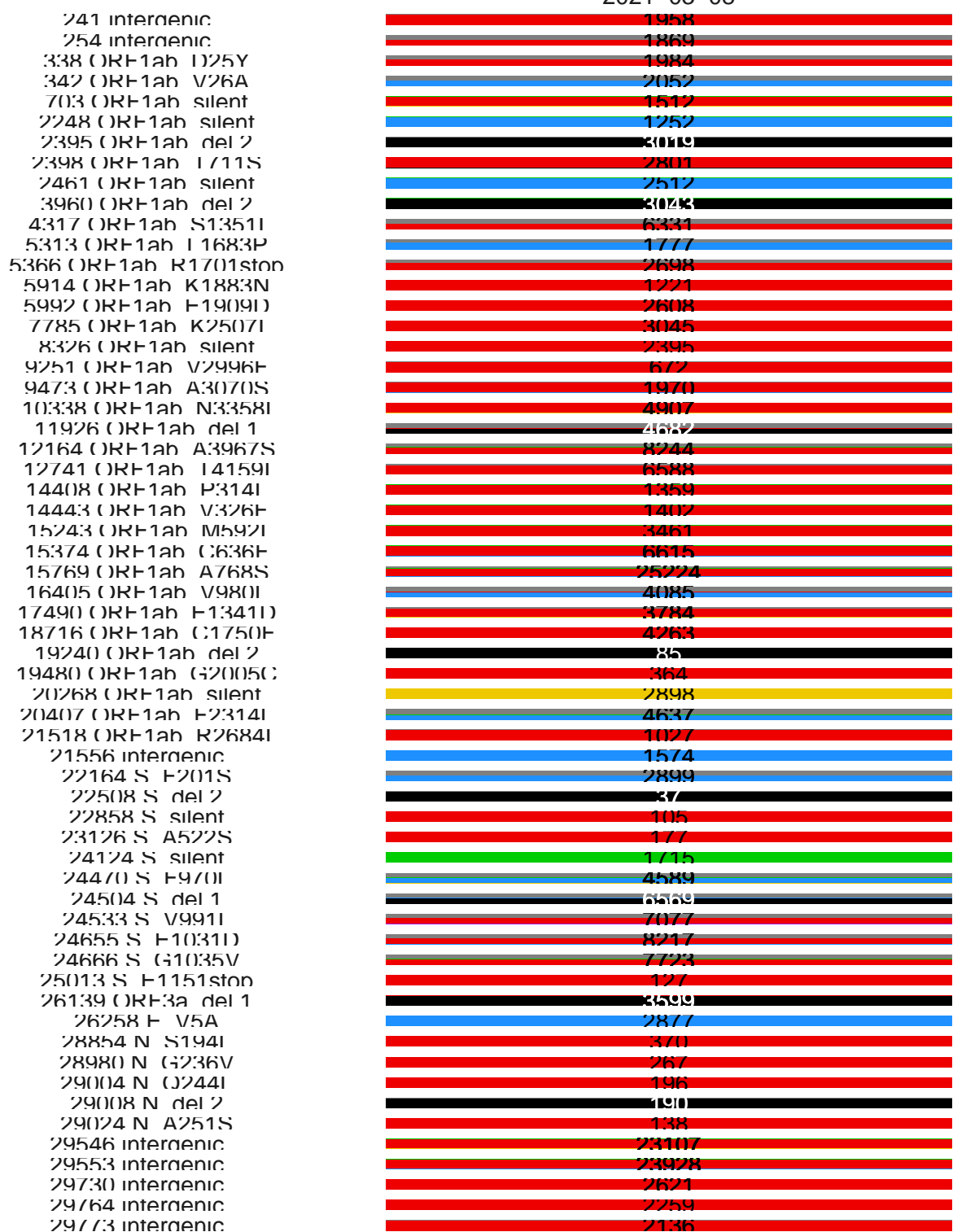
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0969-1	single experiment	NA	Saline	2021-03-08	2.48	NA	58.2%	57.2%

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.

241 intergenic — 1958
254 intergenic — 1869
338 ORF1ab D25Y — 1984
342 ORF1ab V26A — 2052
703 ORF1ab silent — 1512
2248 ORF1ab silent — 1252
2395 ORF1ab del 2 — 3019
2398 ORF1ab I711S — 2801
2461 ORF1ab silent — 2512
3960 ORF1ab del 2 — 3043
4317 ORF1ab S1351I — 6331
5313 ORF1ab L1683P — 1777
5366 ORF1ab R1701stop — 2698
5914 ORF1ab K1883N — 1221
5992 ORF1ab F1909D — 2608
7785 ORF1ab K2507I — 3045
8326 ORF1ab silent — 2395
9251 ORF1ab V2996F — 672
9473 ORF1ab A3070S — 1970
10338 ORF1ab N3358I — 4907
11926 ORF1ab del 1 — 4682
12164 ORF1ab A3967S — 8244
12741 ORF1ab L4159I — 6588
14408 ORF1ab P314L — 1359
14443 ORF1ab V326F — 1402
15243 ORF1ab M592I — 3461
15374 ORF1ab C636F — 6615
15769 ORF1ab A768S — 25224
16405 ORF1ab V980I — 4085
17490 ORF1ab F1341D — 3784
18716 ORF1ab C1750F — 4263
19240 ORF1ab del 2 — 85
19480 ORF1ab G2005C — 364
20268 ORF1ab silent — 2898
20407 ORF1ab F2314I — 4637
21518 ORF1ab R2684I — 1027
21556 intergenic — 1574
22164 S F201S — 2899
22508 S del 2 — 37
22858 S silent — 105
23126 S A522S — 177
24124 S silent — 1715
24470 S F970I — 4589
24504 S del 1 — 6569
24533 S V991I — 7077
24655 S F1031D — 8217
24666 S G1035V — 7723
25013 S F1151stop — 127
26139 ORF3a del 1 — 3599
26258 E V5A — 2877
28854 N S194I — 370
28980 N G236V — 267
29004 N Q244I — 196
29008 N del 2 — 190
29024 N A251S — 138
29546 intergenic — 23107
29553 intergenic — 23928
29730 intergenic — 2621
29764 intergenic — 2259
29773 intergenic — 2136

VSP0969−1

Base change

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

3

Analyses of individual experiments and composite results

VSP0969-1 | 2021-03-08 | Saline | UPHS-0037 | 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.
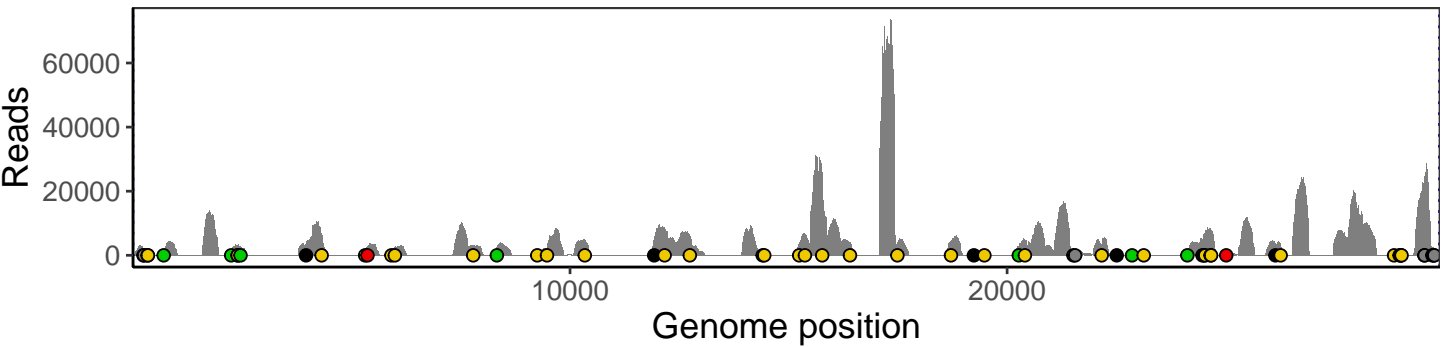
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
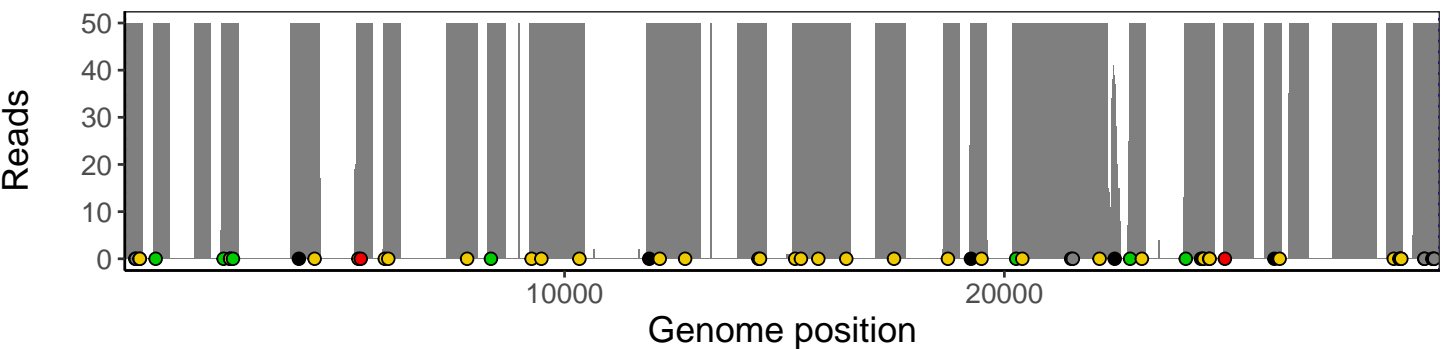
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
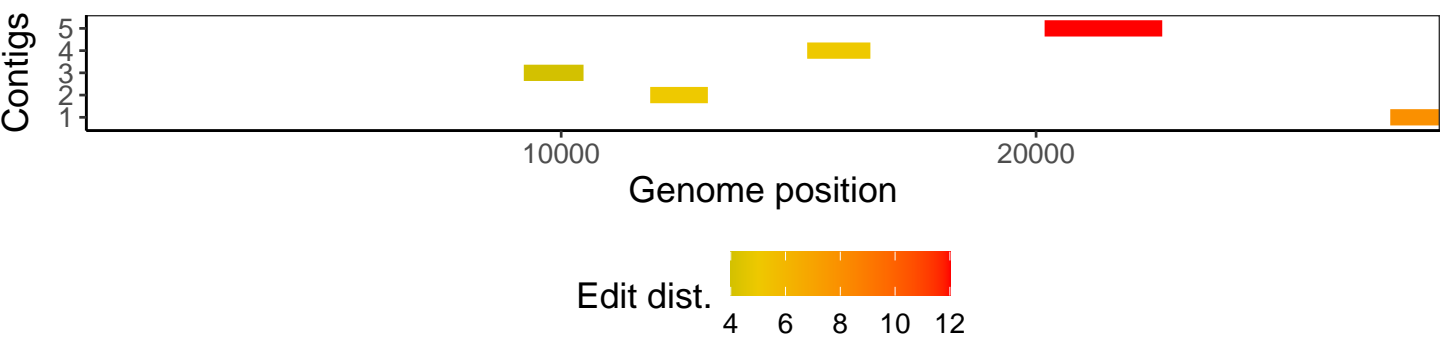
Software environment

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