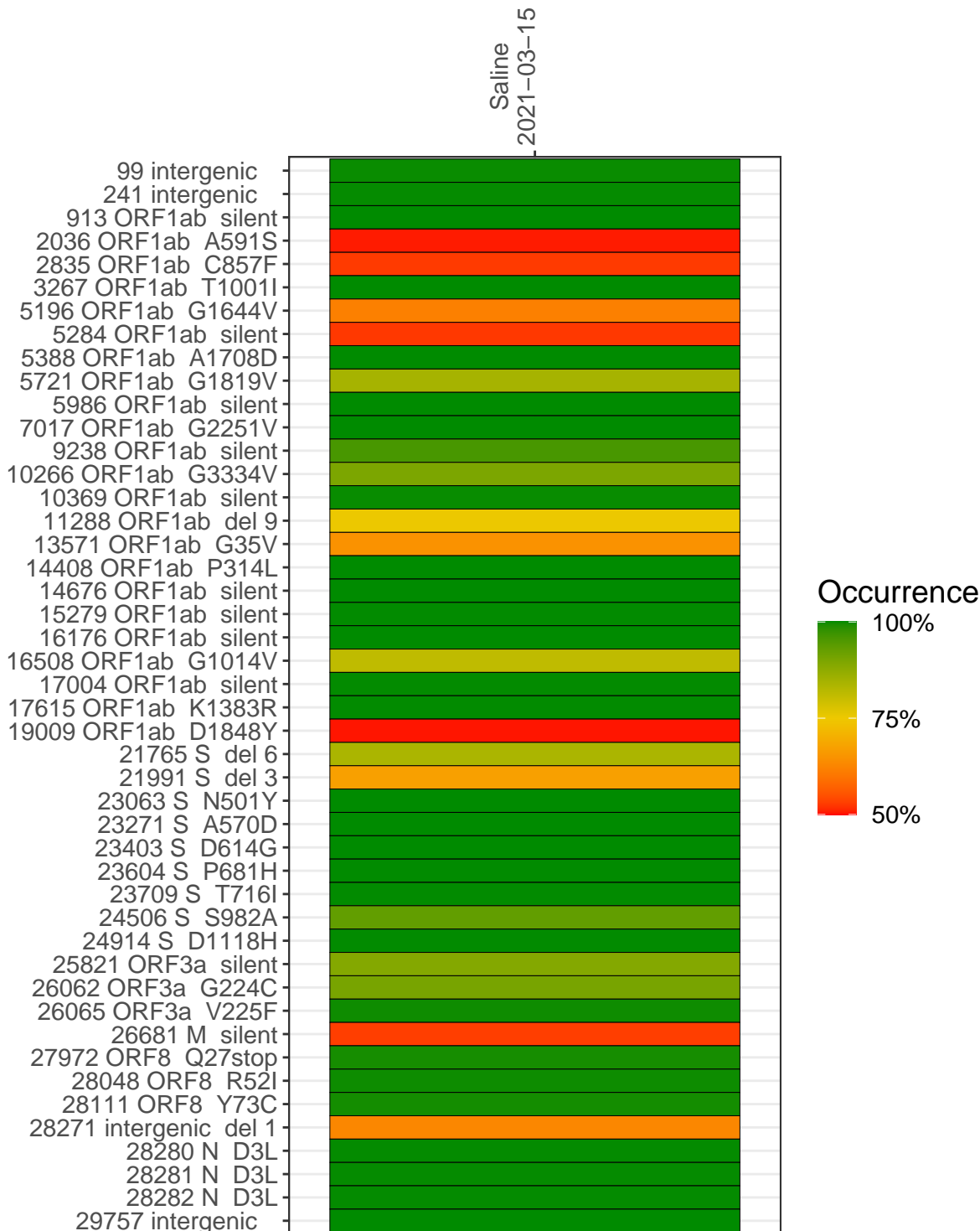# COVID-19 subject HUP Q-0043

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.
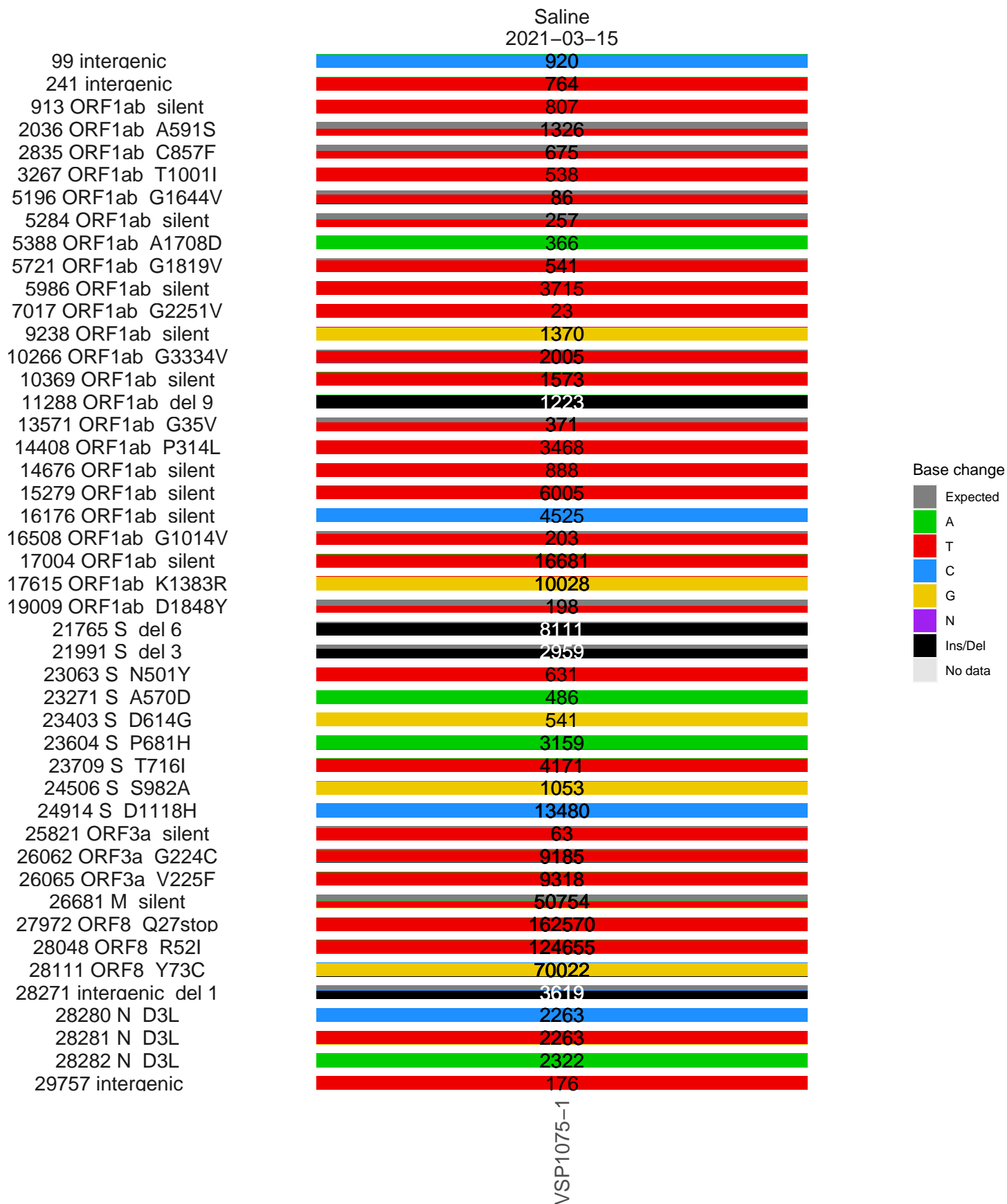
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1075-1	single experiment	NA	Saline	2021-03-15	24.61	B.1.1.7	97.4%	97.3%

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
2021-03-15

Position	Count
99 intergenic	920
241 intergenic	764
913 ORF1ab silent	807
2036 ORF1ab A591S	1326
2835 ORF1ab C857F	675
3267 ORF1ab T1001I	538
5196 ORF1ab G1644V	86
5284 ORF1ab silent	257
5388 ORF1ab A1708D	366
5721 ORF1ab G1819V	541
5986 ORF1ab silent	3715
7017 ORF1ab G2251V	23
9238 ORF1ab silent	1370
10266 ORF1ab G3334V	2005
10369 ORF1ab silent	1573
11288 ORF1ab del 9	1223
13571 ORF1ab G35V	371
14408 ORF1ab P314L	3468
14676 ORF1ab silent	888
15279 ORF1ab silent	6005
16176 ORF1ab silent	4525
16508 ORF1ab G1014V	203
17004 ORF1ab silent	16681
17615 ORF1ab K1383R	10028
19009 ORF1ab D1848Y	198
21765 S del 6	8111
21991 S del 3	2959
23063 S N501Y	631
23271 S A570D	486
23403 S D614G	541
23604 S P681H	3159
23709 S T716I	4171
24506 S S982A	1053
24914 S D1118H	13480
25821 ORF3a silent	63
26062 ORF3a G224C	9185
26065 ORF3a V225F	9318
26681 M silent	50754
27972 ORF8 Q27stop	162570
28048 ORF8 R52I	124655
28111 ORF8 Y73C	70022
28271 intergenic del 1	3619
28280 N D3L	2263
28281 N D3L	2263
28282 N D3L	2322
29757 intergenic	176

VSP1075-1

Base change
- Expected (gray)
- A (green)
- T (red)
- C (blue)
- G (yellow)
- N (purple)
- Ins/Del (black)
- No data (light gray)

3

Analyses of individual experiments and composite results

**VSP1075-1 | 2021-03-15 | Saline | HUP Q-0043 | 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.
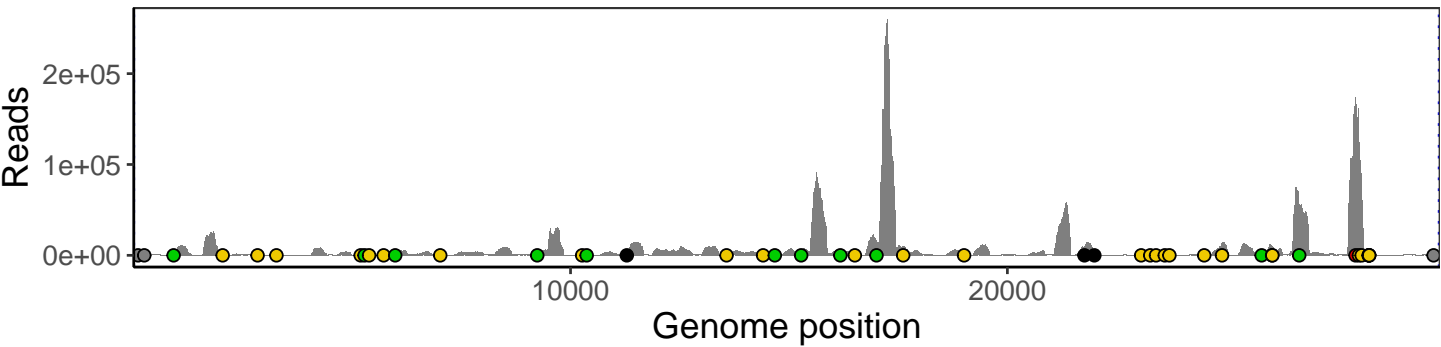
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
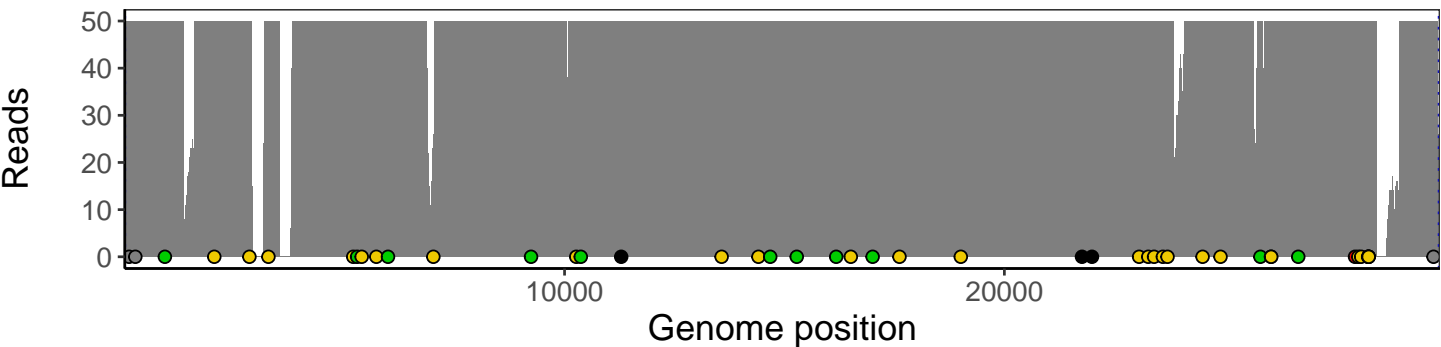
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
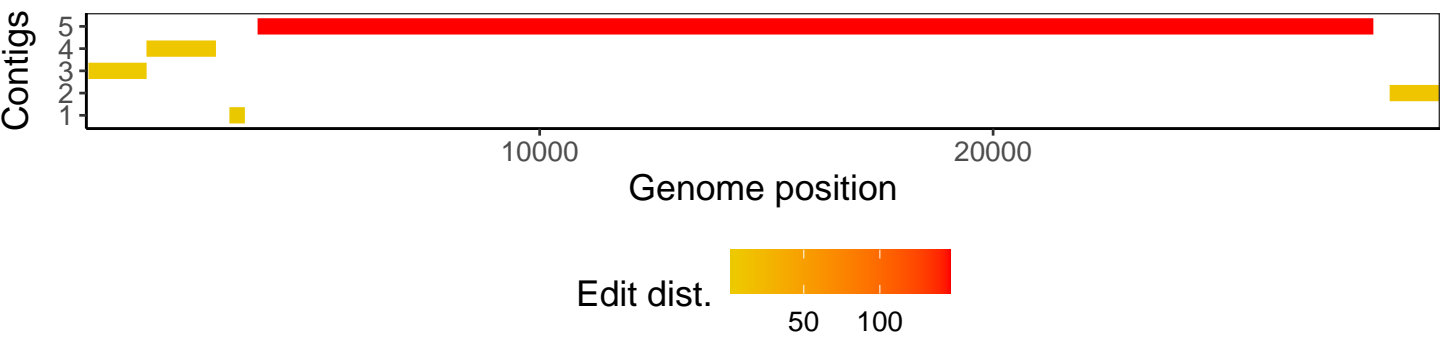
Software environment

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