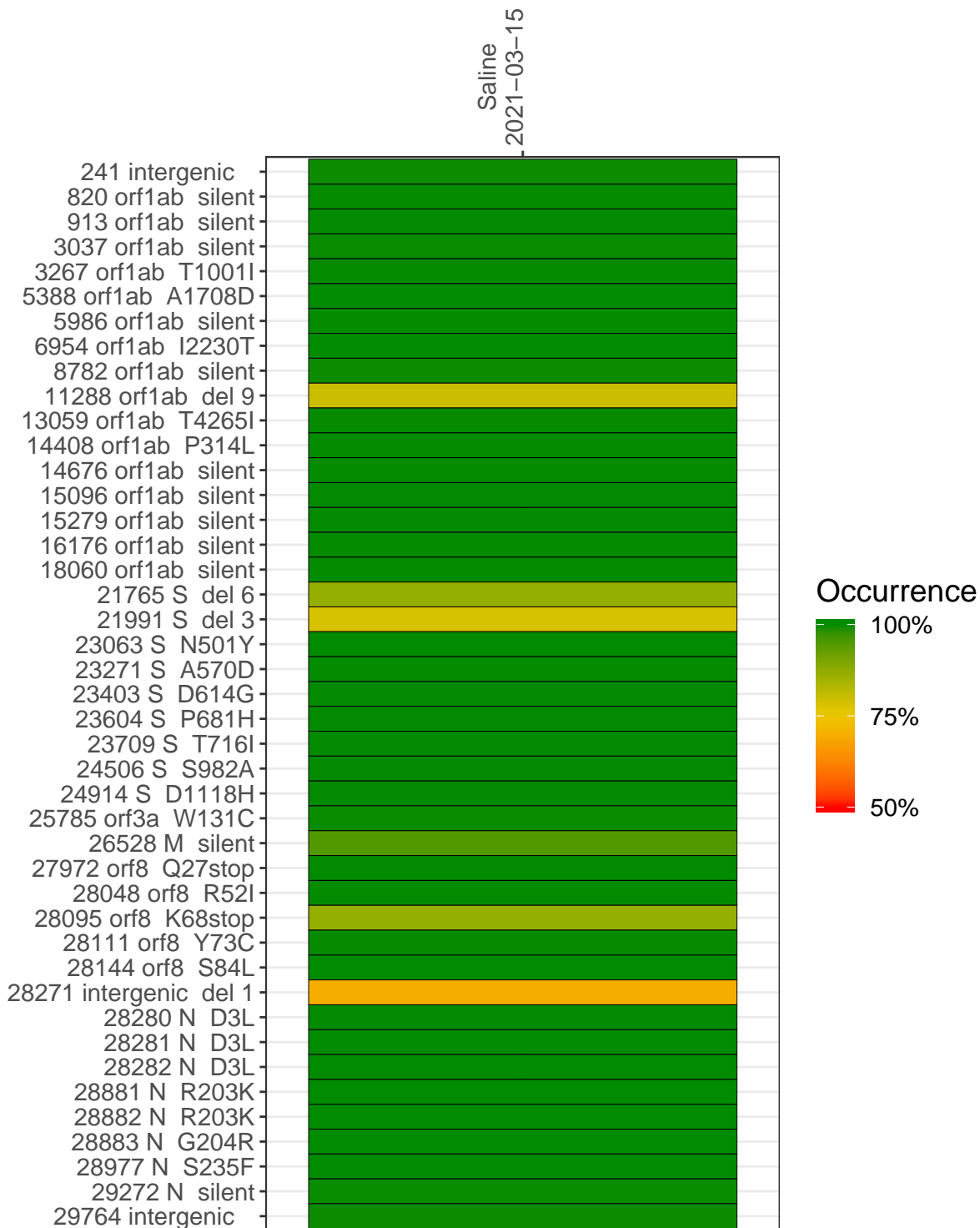# COVID-19 subject HUP Q-0049

2021-04-01

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.
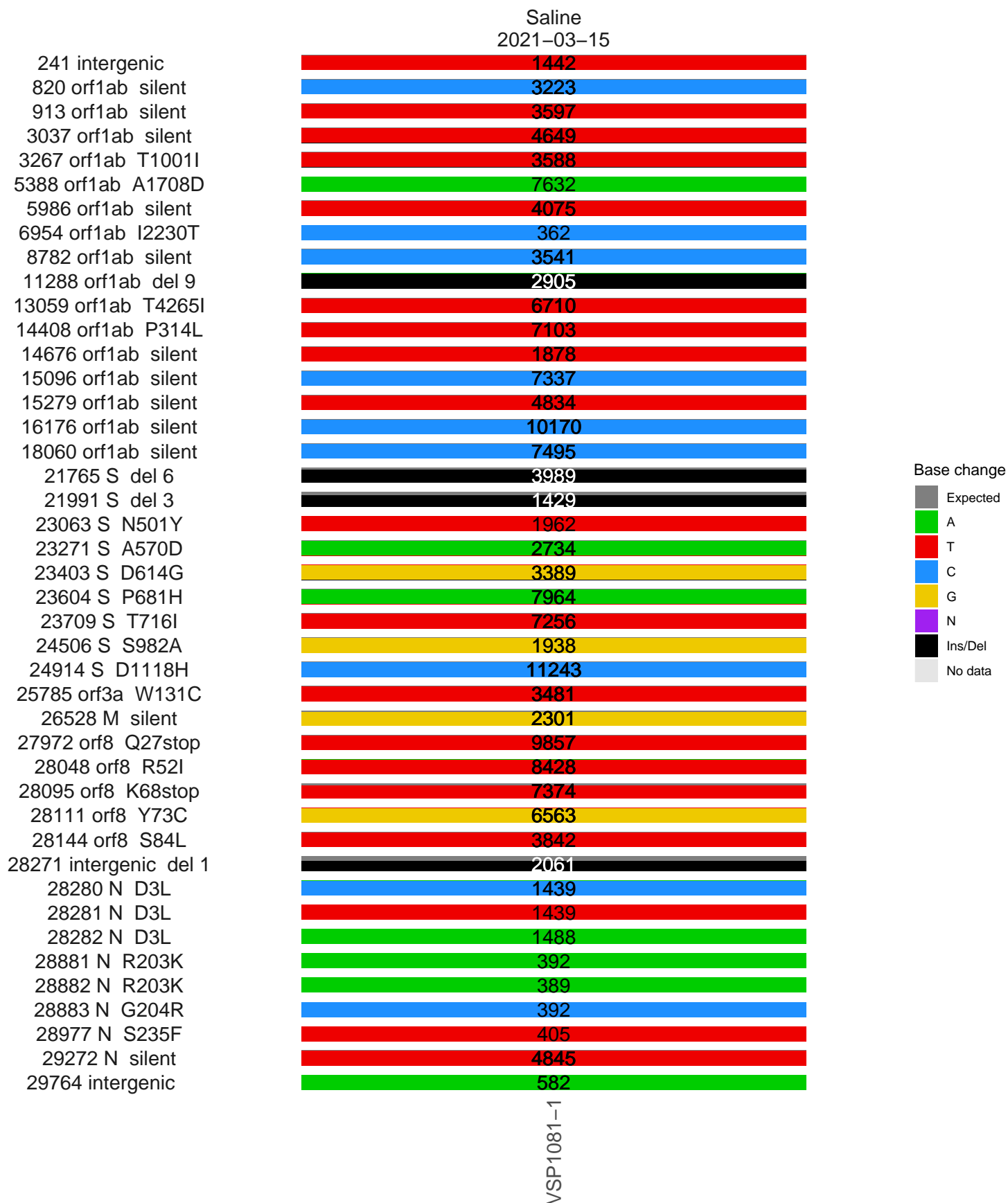
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1081-1	single experiment	NA	Saline	2021-03-15	29.85	B.1.1.7	99.9%	99.9%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/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
2021−03−15

Position / Gene / Change	Value
241 intergenic	1442
820 orf1ab silent	3223
913 orf1ab silent	3597
3037 orf1ab silent	4649
3267 orf1ab T1001I	3588
5388 orf1ab A1708D	7632
5986 orf1ab silent	4075
6954 orf1ab I2230T	362
8782 orf1ab silent	3541
11288 orf1ab del 9	2905
13059 orf1ab T4265I	6710
14408 orf1ab P314L	7103
14676 orf1ab silent	1878
15096 orf1ab silent	7337
15279 orf1ab silent	4834
16176 orf1ab silent	10170
18060 orf1ab silent	7495
21765 S del 6	3989
21991 S del 3	1429
23063 S N501Y	1962
23271 S A570D	2734
23403 S D614G	3389
23604 S P681H	7964
23709 S T716I	7256
24506 S S982A	1938
24914 S D1118H	11243
25785 orf3a W131C	3481
26528 M silent	2301
27972 orf8 Q27stop	9857
28048 orf8 R52I	8428
28095 orf8 K68stop	7374
28111 orf8 Y73C	6563
28144 orf8 S84L	3842
28271 intergenic del 1	2061
28280 N D3L	1439
28281 N D3L	1439
28282 N D3L	1488
28881 N R203K	392
28882 N R203K	389
28883 N G204R	392
28977 N S235F	405
29272 N silent	4845
29764 intergenic	582

VSP1081−1

Base change

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

3

Analyses of individual experiments and composite results

## VSP1081-1 | 2021-03-15 | Saline | HUP Q-0049 | 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.
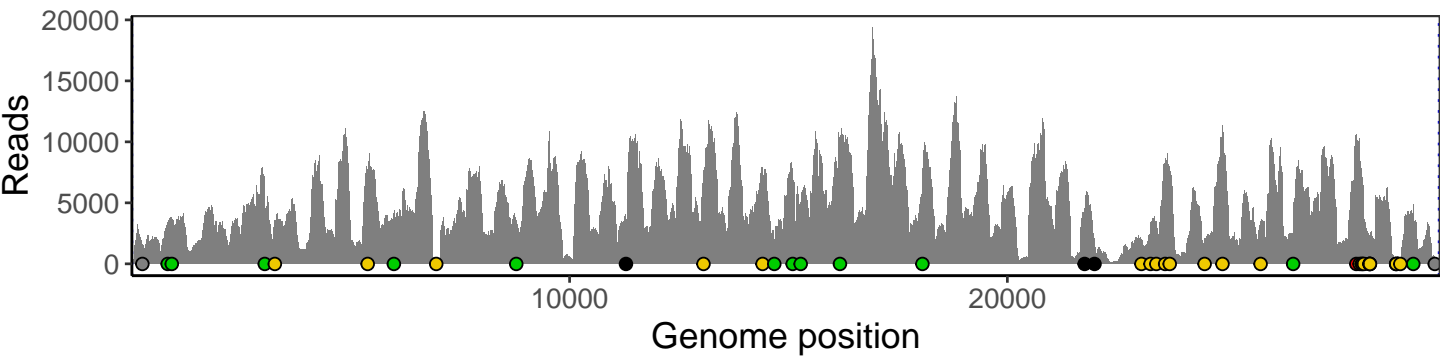
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
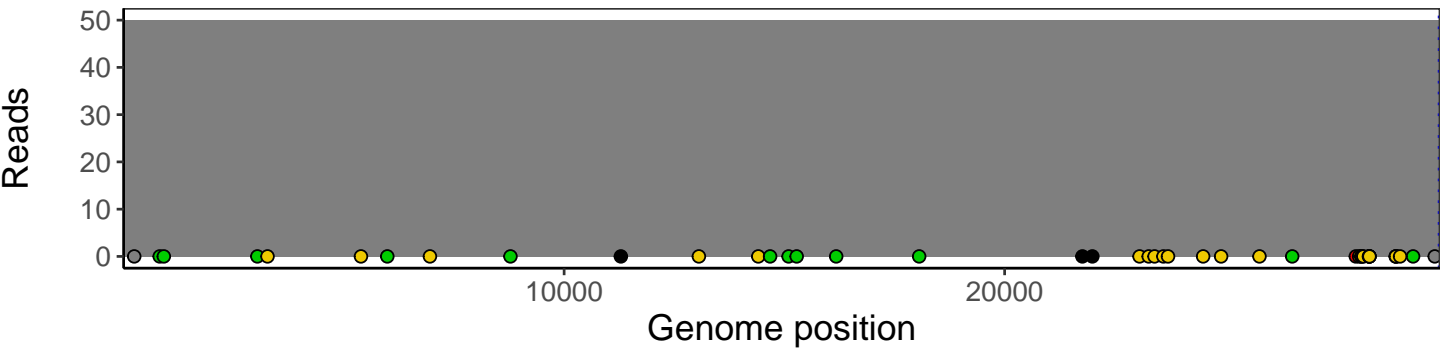
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
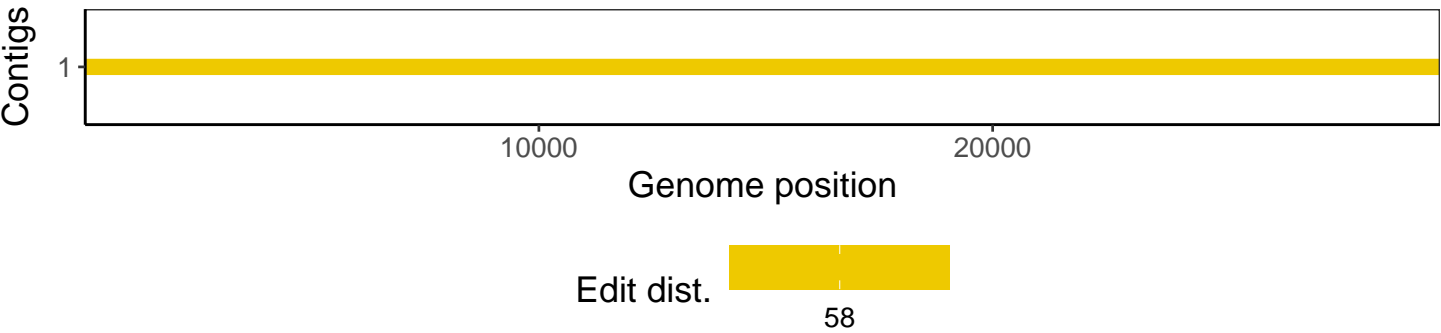
Software environment

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