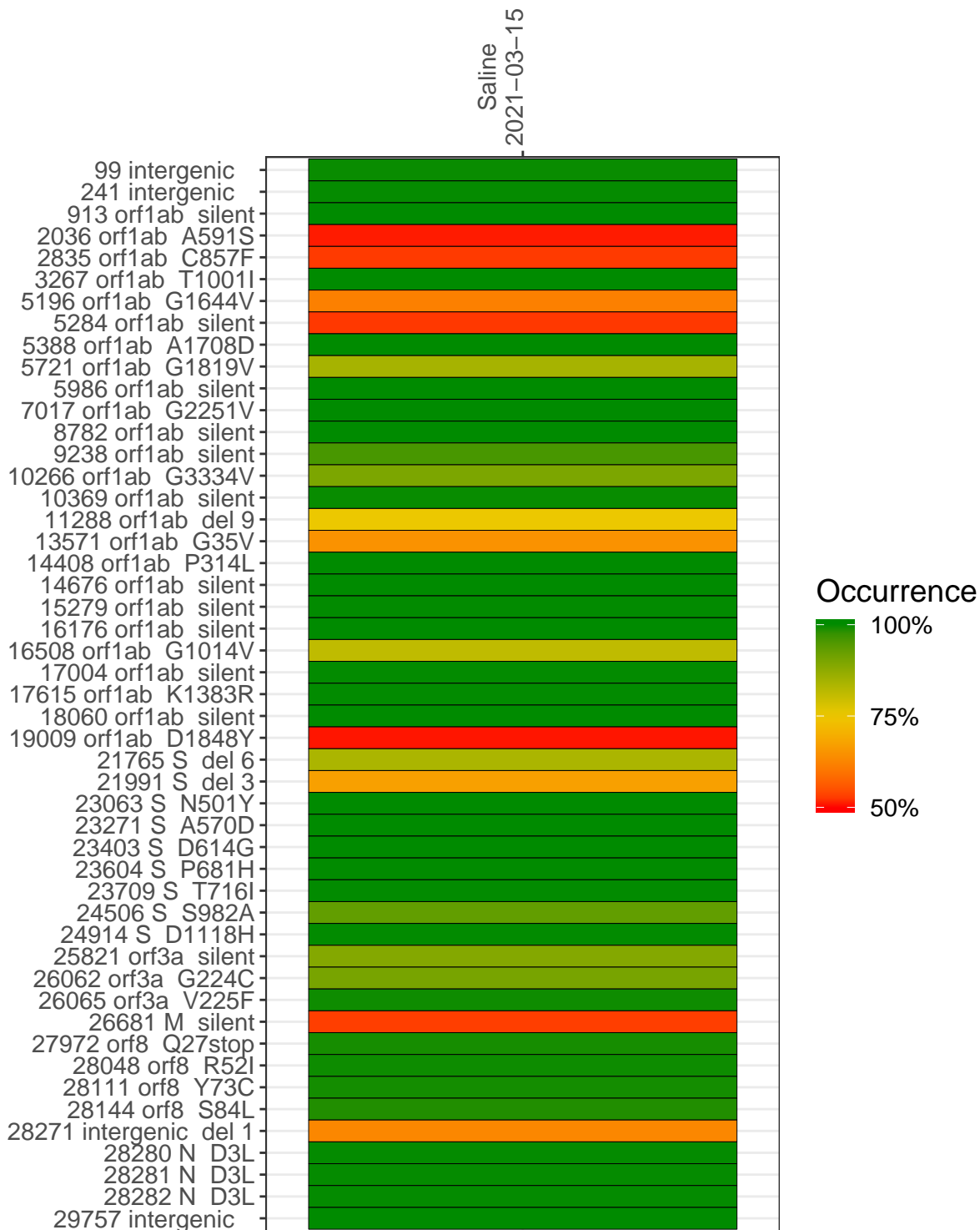# COVID-19 subject HUP Q-0043

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.
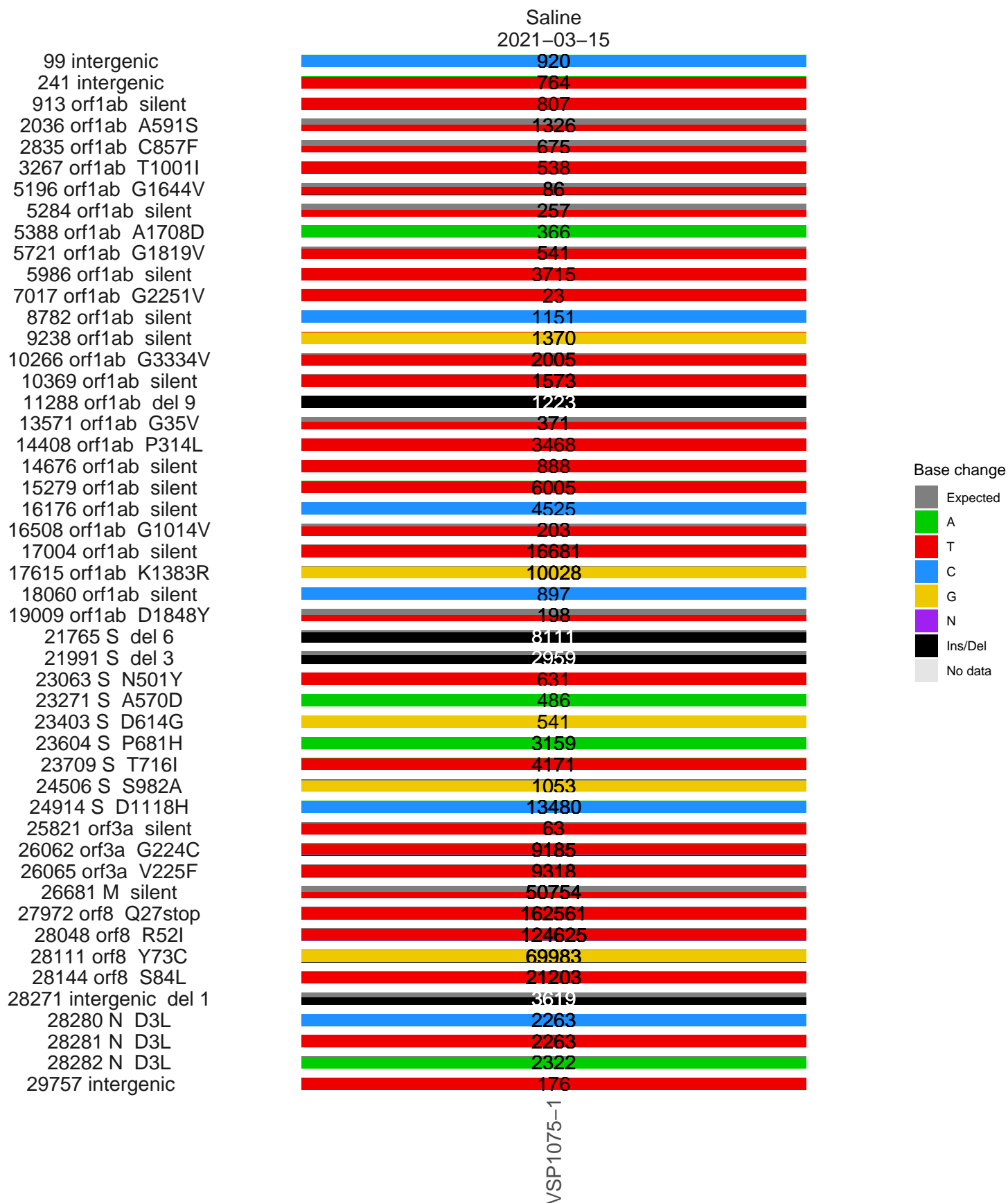
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1075-1	single experiment	NA	Saline	2021-03-15	24.61	B.1.1.7	97.3%	97.2%

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

Variant	Value	Base change
99 intergenic	920	C
241 intergenic	764	T
913 orf1ab silent	807	T
2036 orf1ab A591S	1326	Expected
2835 orf1ab C857F	675	Expected
3267 orf1ab T1001I	538	T
5196 orf1ab G1644V	86	Expected
5284 orf1ab silent	257	Expected
5388 orf1ab A1708D	366	A
5721 orf1ab G1819V	541	Expected
5986 orf1ab silent	3715	T
7017 orf1ab G2251V	23	T
8782 orf1ab silent	1151	C
9238 orf1ab silent	1370	G
10266 orf1ab G3334V	2005	T
10369 orf1ab silent	1573	T
11288 orf1ab del 9	1223	Ins/Del
13571 orf1ab G35V	371	Expected
14408 orf1ab P314L	3468	T
14676 orf1ab silent	888	T
15279 orf1ab silent	6005	T
16176 orf1ab silent	4525	C
16508 orf1ab G1014V	203	T
17004 orf1ab silent	16681	T
17615 orf1ab K1383R	10028	G
18060 orf1ab silent	897	C
19009 orf1ab D1848Y	198	Expected/T
21765 S del 6	8111	Ins/Del
21991 S del 3	2959	Ins/Del
23063 S N501Y	631	T
23271 S A570D	486	A
23403 S D614G	541	G
23604 S P681H	3159	A
23709 S T716I	4171	T
24506 S S982A	1053	G
24914 S D1118H	13480	C
25821 orf3a silent	63	T
26062 orf3a G224C	9185	T
26065 orf3a V225F	9318	T
26681 M silent	50754	Expected
27972 orf8 Q27stop	162561	T
28048 orf8 R52I	124625	T
28111 orf8 Y73C	69983	G
28144 orf8 S84L	21203	T
28271 intergenic del 1	3619	Ins/Del
28280 N D3L	2263	C
28281 N D3L	2263	T
28282 N D3L	2322	A
29757 intergenic	176	T

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

VSP1075−1

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.
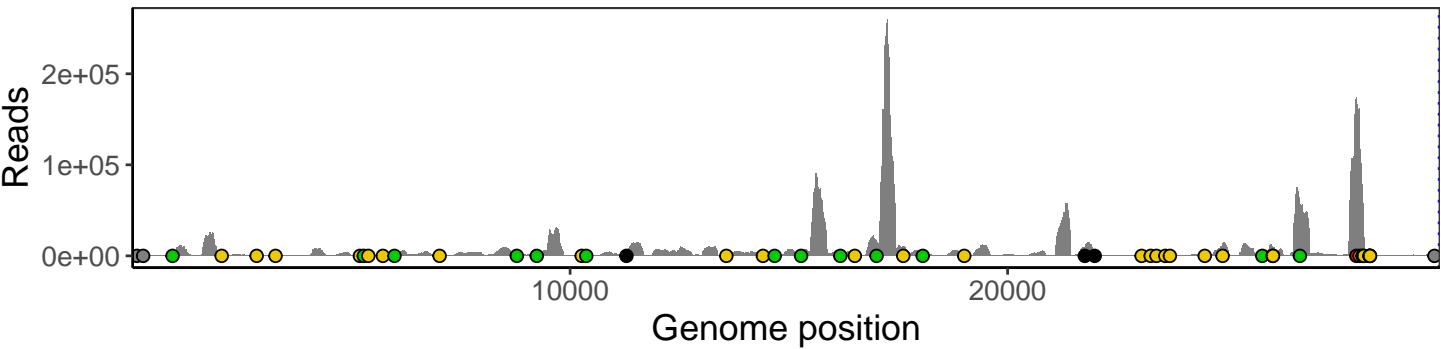
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
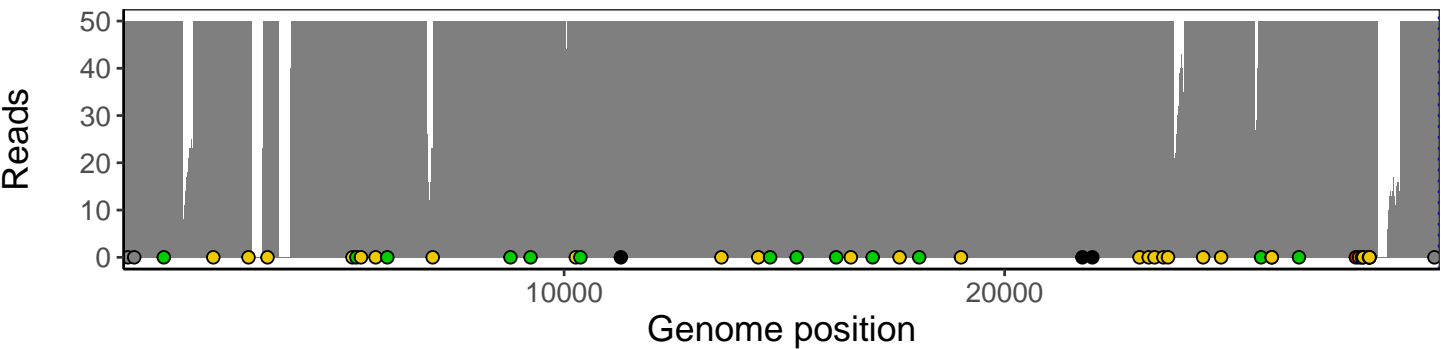
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
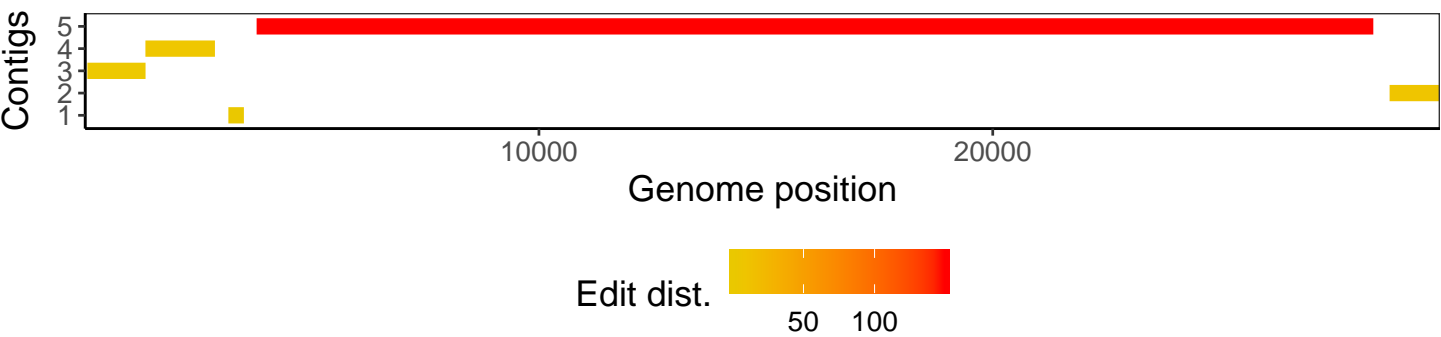
Software environment

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	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