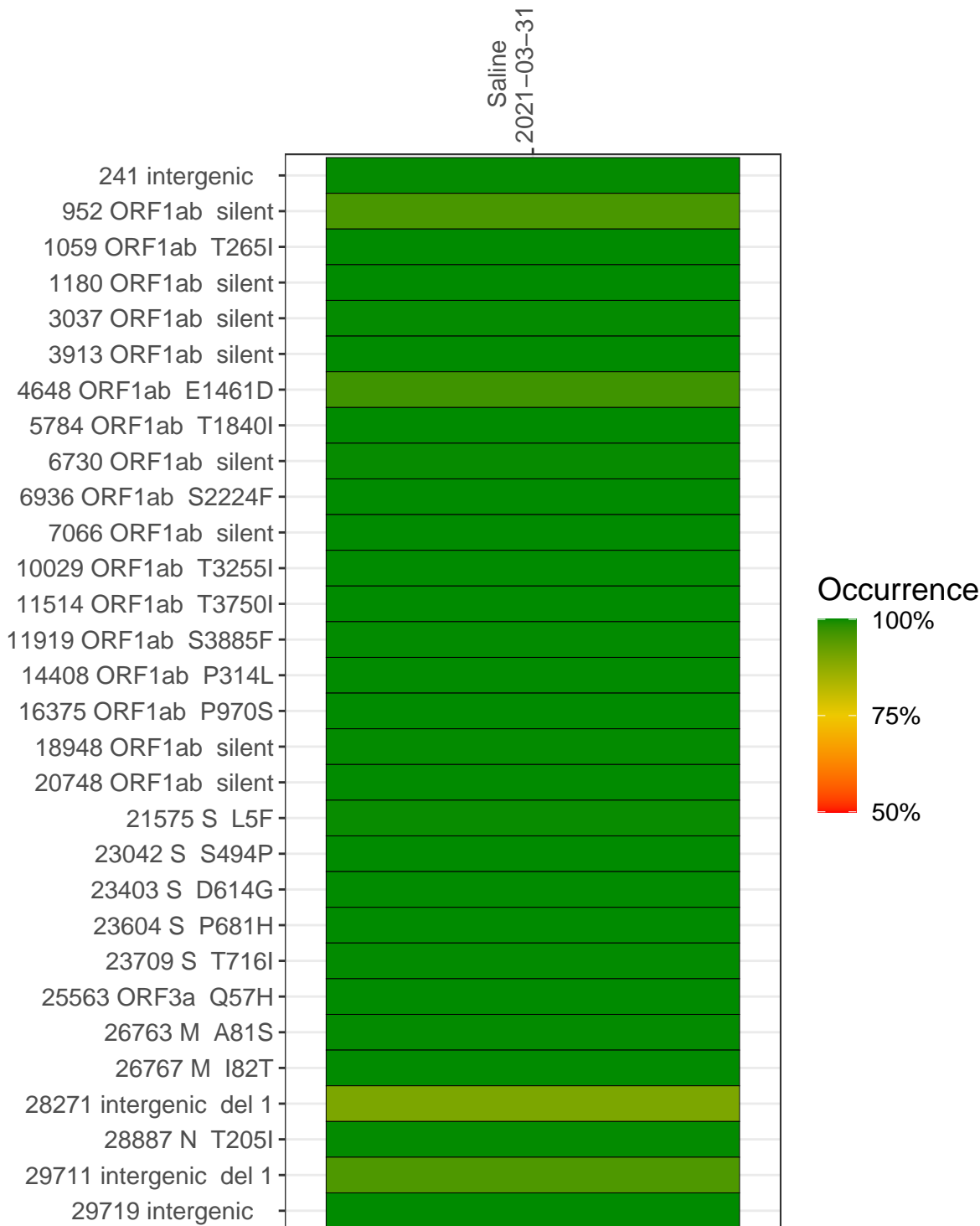# COVID-19 subject UPHS-0595

2021-05-05

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.
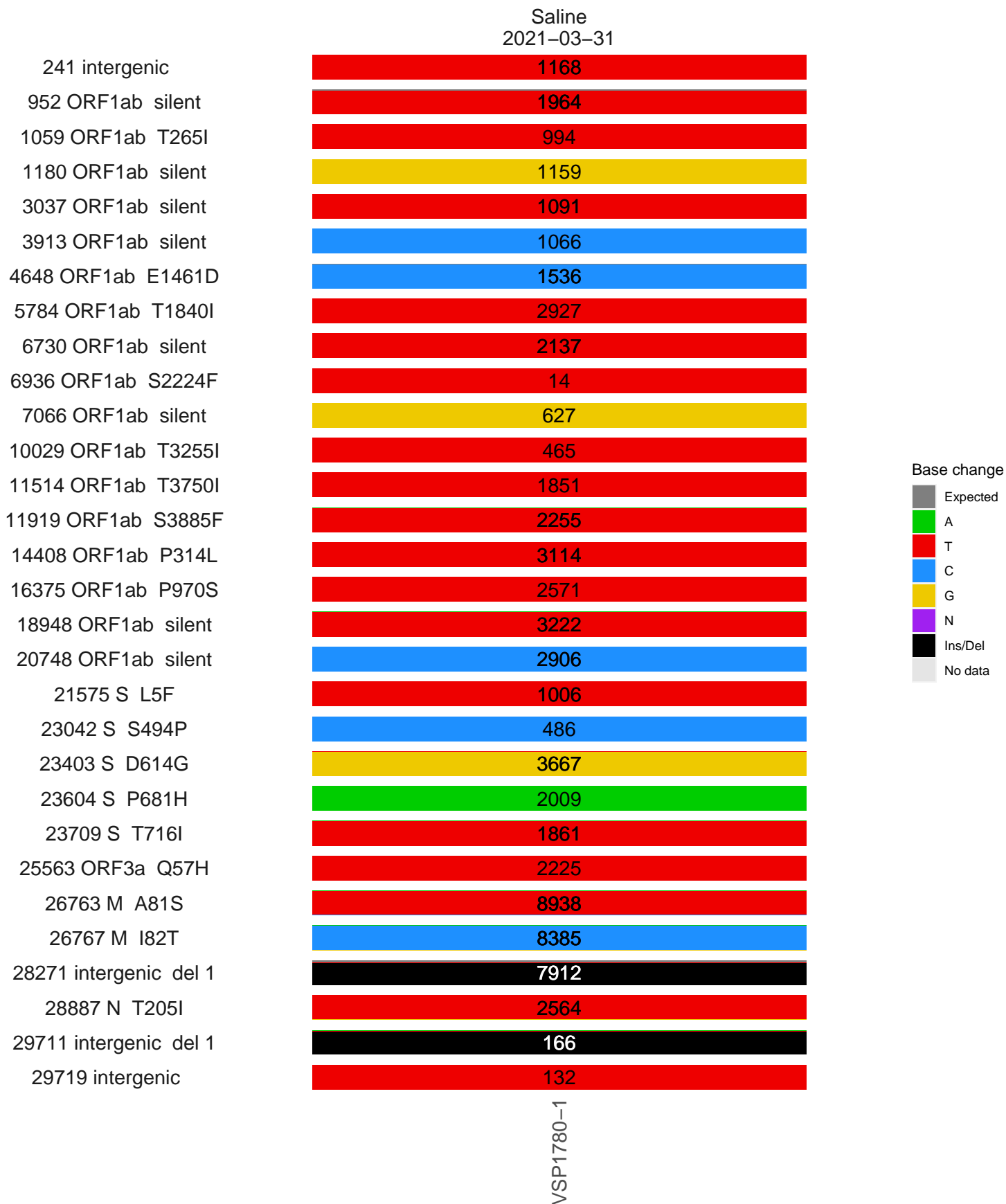
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1780-1	single experiment	NA	Saline	2021-03-31	29.82	B.1.575	99.8%	99.8%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/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
2021−03−31

Position / Gene / Mutation	Count	Base change
241 intergenic	1168	T
952 ORF1ab silent	1964	T
1059 ORF1ab T265I	994	T
1180 ORF1ab silent	1159	G
3037 ORF1ab silent	1091	T
3913 ORF1ab silent	1066	C
4648 ORF1ab E1461D	1536	C
5784 ORF1ab T1840I	2927	T
6730 ORF1ab silent	2137	T
6936 ORF1ab S2224F	14	T
7066 ORF1ab silent	627	G
10029 ORF1ab T3255I	465	T
11514 ORF1ab T3750I	1851	T
11919 ORF1ab S3885F	2255	T
14408 ORF1ab P314L	3114	T
16375 ORF1ab P970S	2571	T
18948 ORF1ab silent	3222	T
20748 ORF1ab silent	2906	C
21575 S L5F	1006	T
23042 S S494P	486	C
23403 S D614G	3667	G
23604 S P681H	2009	A
23709 S T716I	1861	T
25563 ORF3a Q57H	2225	T
26763 M A81S	8938	T
26767 M I82T	8385	C
28271 intergenic del 1	7912	Ins/Del
28887 N T205I	2564	T
29711 intergenic del 1	166	Ins/Del
29719 intergenic	132	T

Base change

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

VSP1780−1

3

Analyses of individual experiments and composite results

VSP1780-1 | 2021-03-31 | Saline | UPHS-0595 | 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.
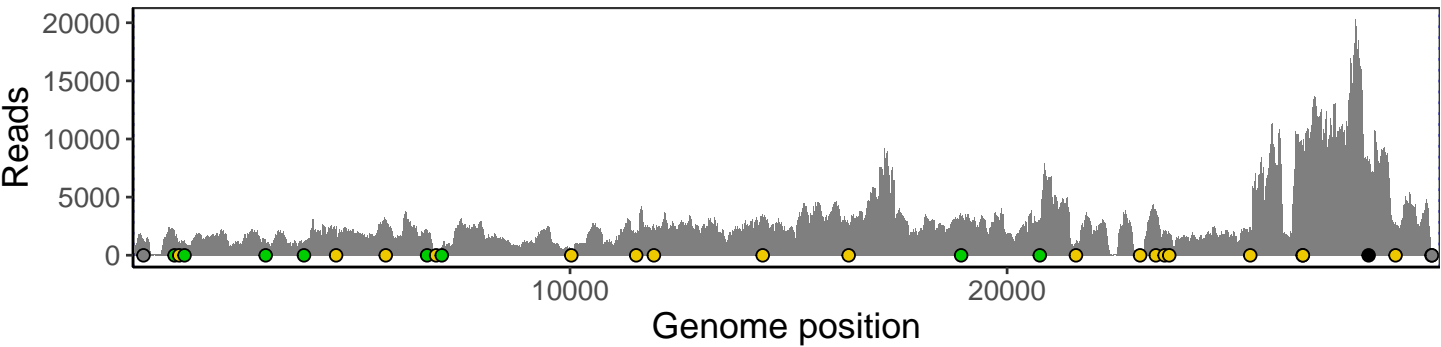
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
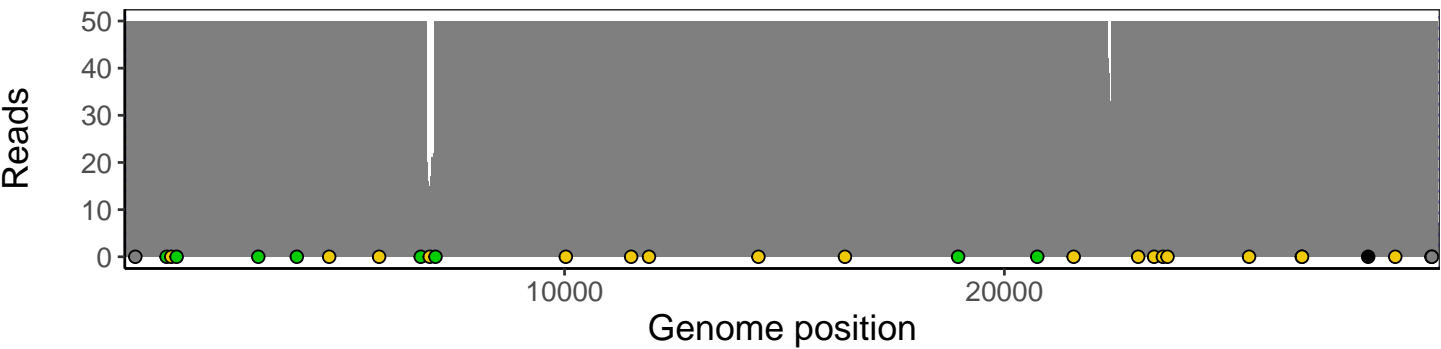
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
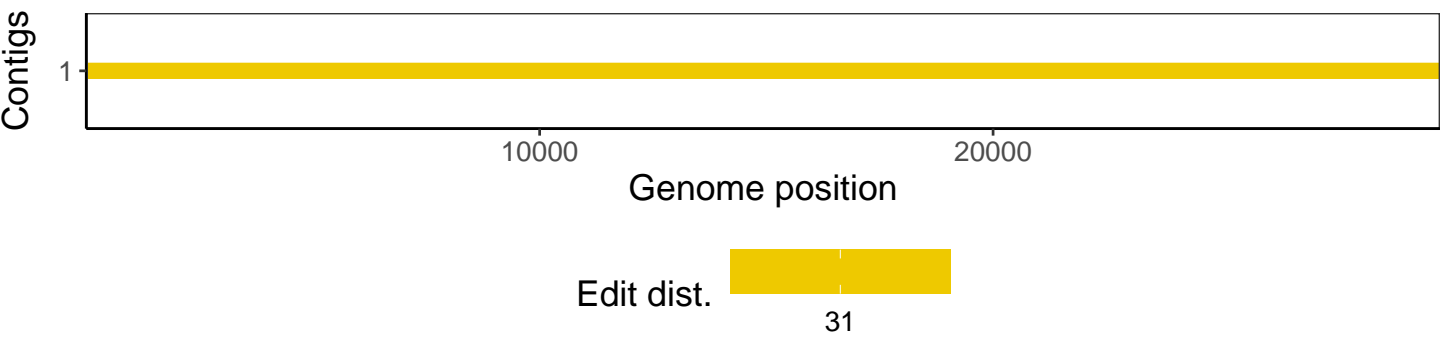
Software environment

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