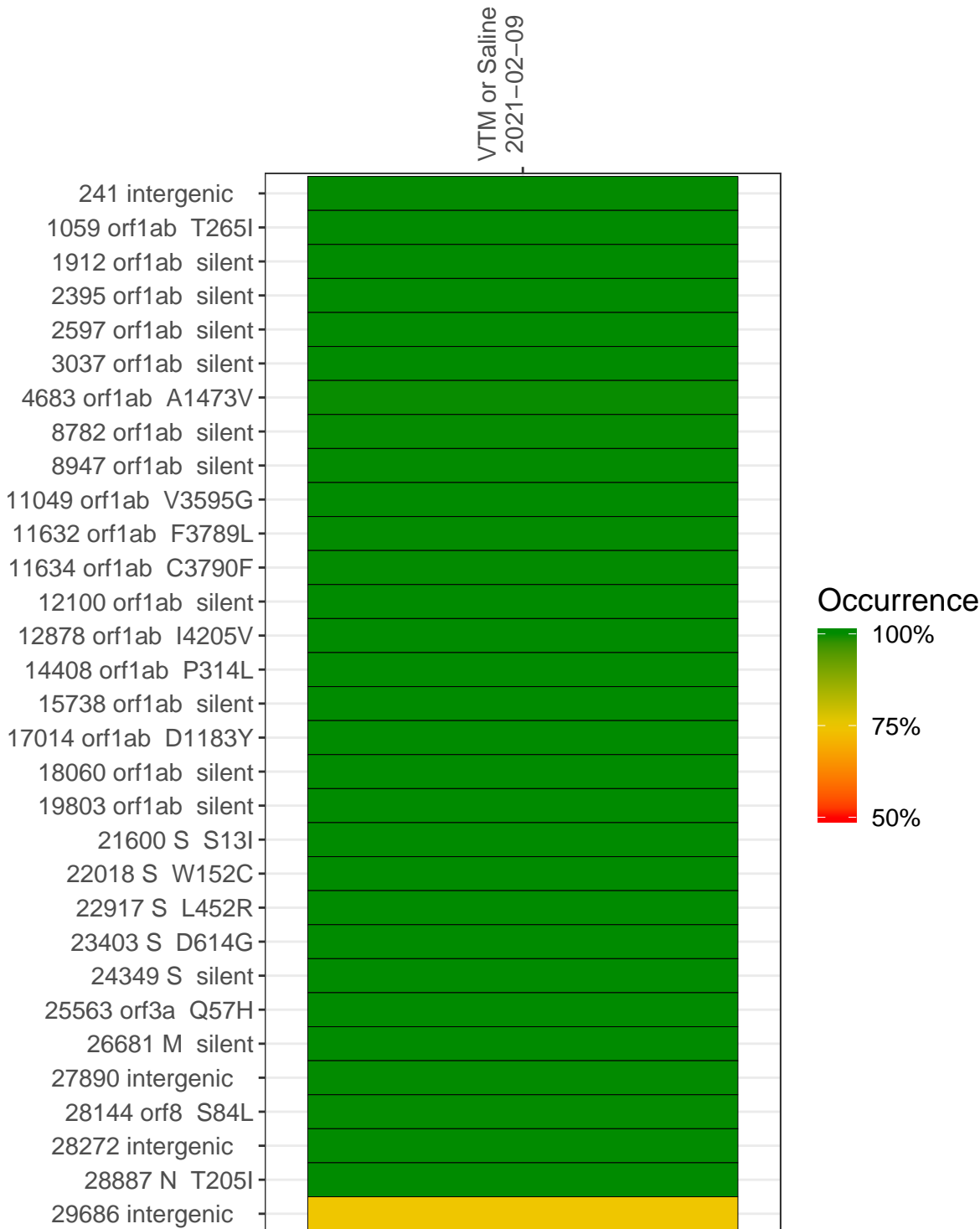# COVID-19 subject H2102030496

2021-03-01

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.
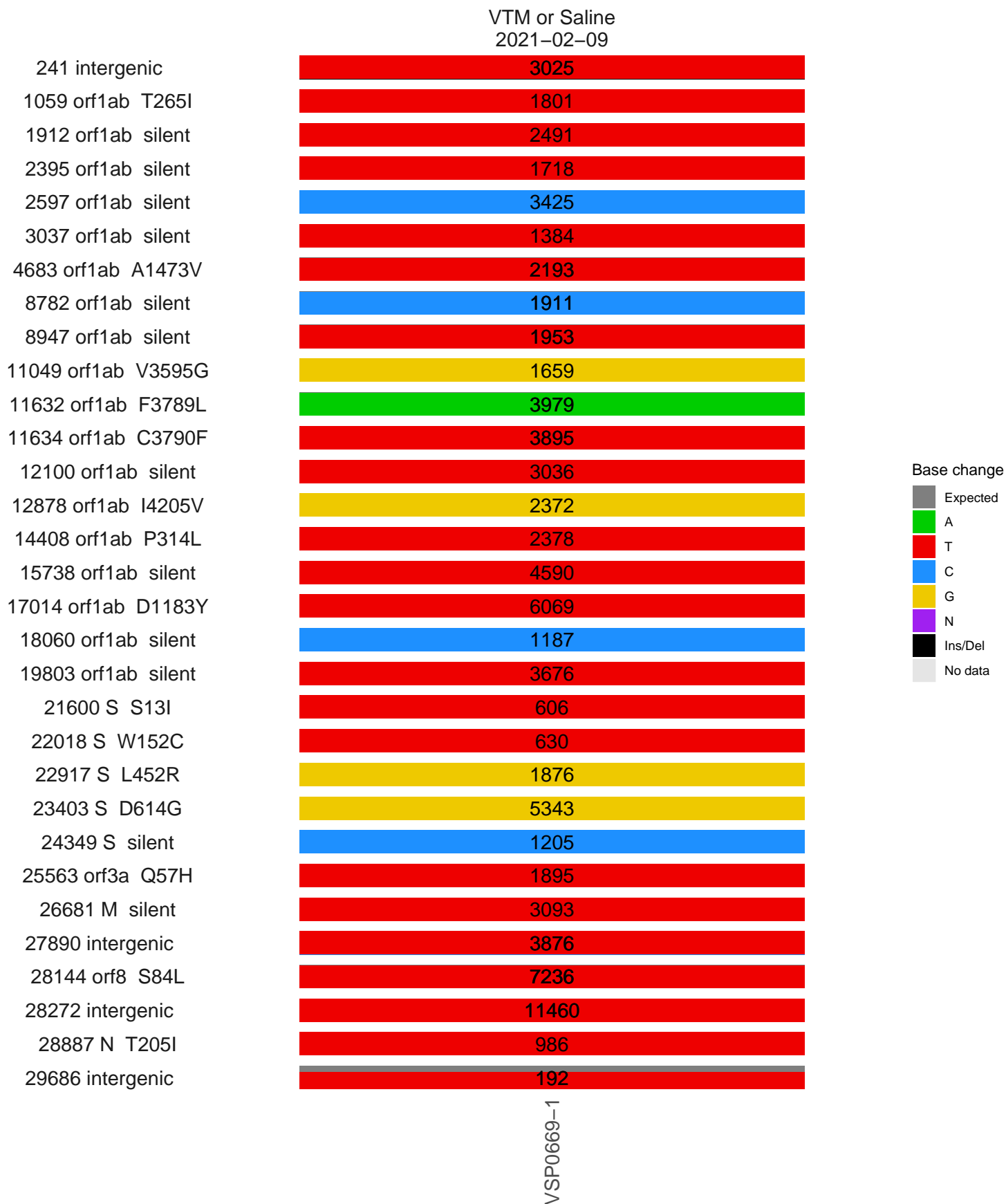
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0669-1	single experiment	NA	VTM or Saline	2021-02-09	21.76	B.1.429	99.0%	99.0%

Variants shared across samples

The heat map below shows how variants (reference genome 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.

VTM or Saline
2021–02–09

Position	Value
241 intergenic	3025
1059 orf1ab T265I	1801
1912 orf1ab silent	2491
2395 orf1ab silent	1718
2597 orf1ab silent	3425
3037 orf1ab silent	1384
4683 orf1ab A1473V	2193
8782 orf1ab silent	1911
8947 orf1ab silent	1953
11049 orf1ab V3595G	1659
11632 orf1ab F3789L	3979
11634 orf1ab C3790F	3895
12100 orf1ab silent	3036
12878 orf1ab I4205V	2372
14408 orf1ab P314L	2378
15738 orf1ab silent	4590
17014 orf1ab D1183Y	6069
18060 orf1ab silent	1187
19803 orf1ab silent	3676
21600 S S13I	606
22018 S W152C	630
22917 S L452R	1876
23403 S D614G	5343
24349 S silent	1205
25563 orf3a Q57H	1895
26681 M silent	3093
27890 intergenic	3876
28144 orf8 S84L	7236
28272 intergenic	11460
28887 N T205I	986
29686 intergenic	192

VSP0669–1

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

3

Analyses of individual experiments and composite results

VSP0669-1 | 2021-02-09 | VTM or Saline | H2102030496 | 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.
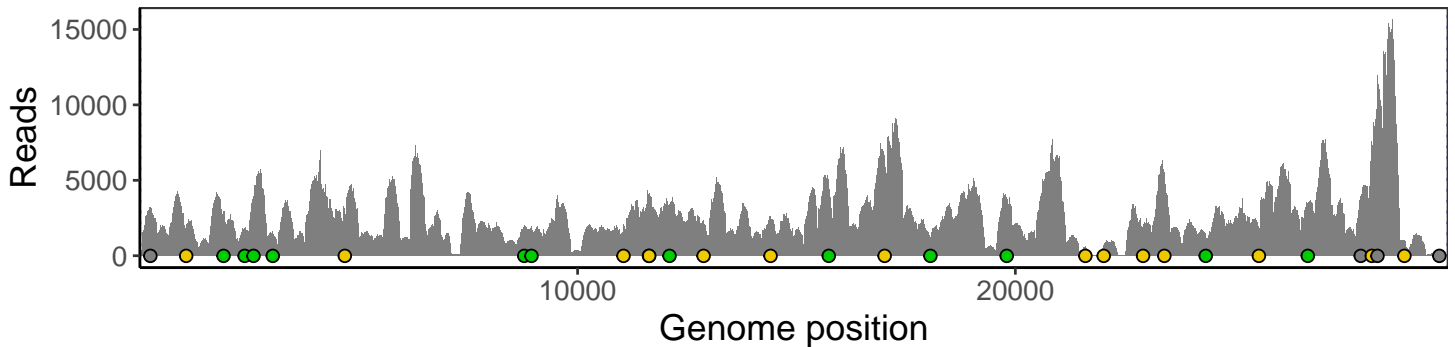
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
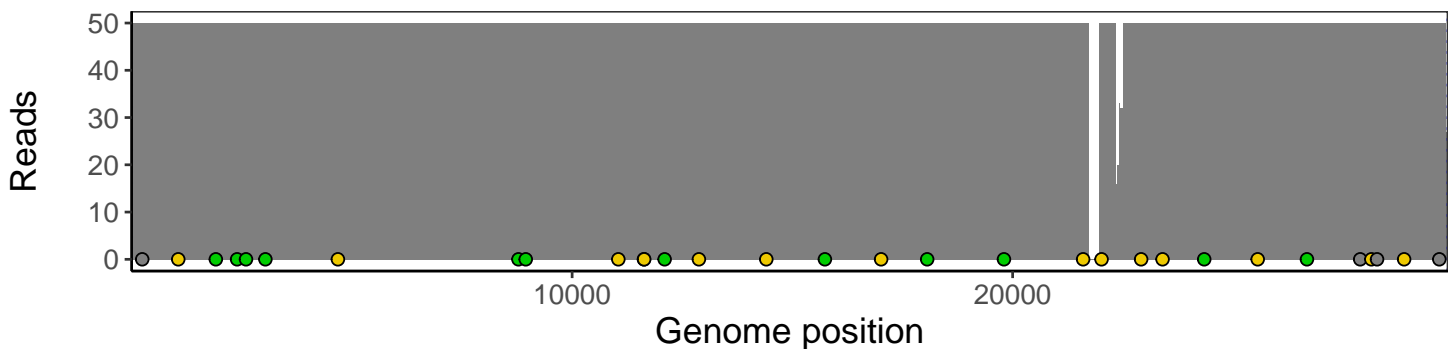
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
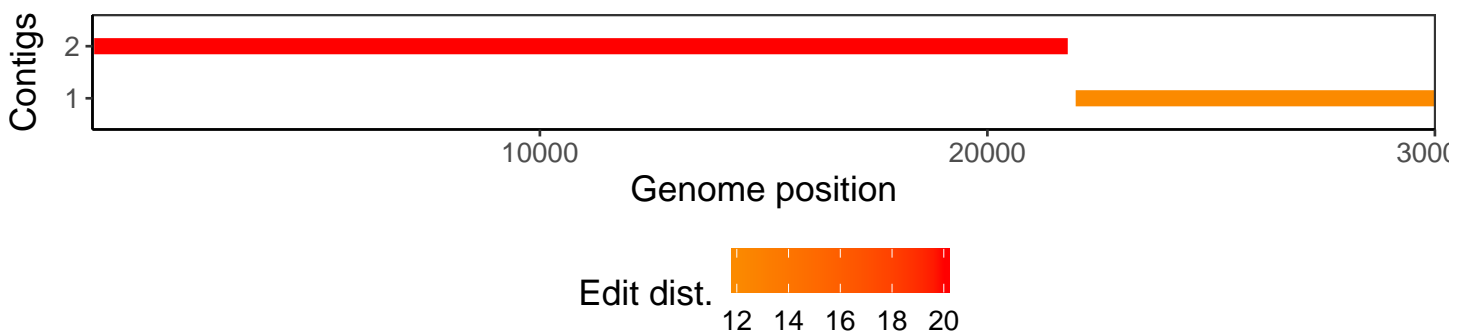
Software environment

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