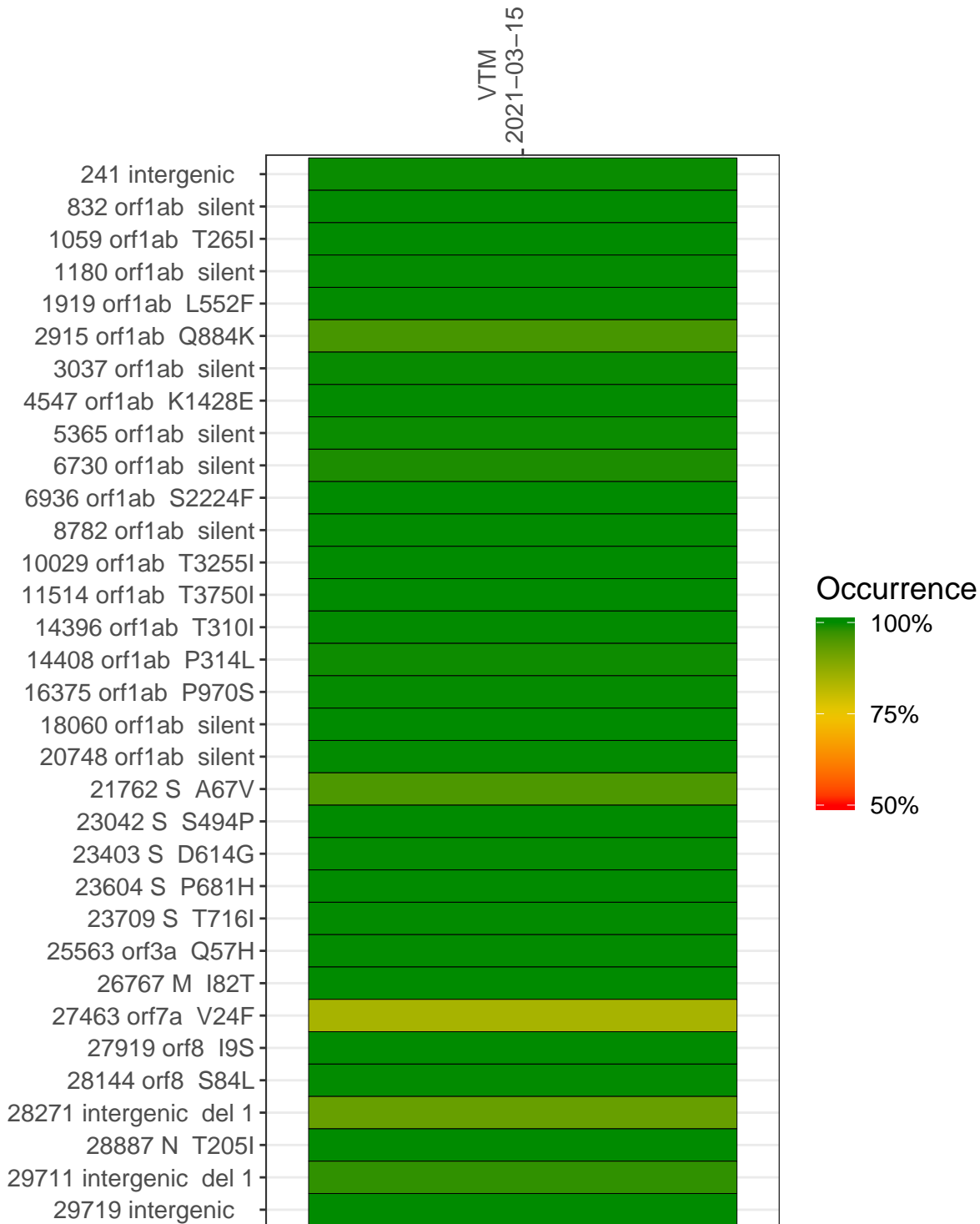# COVID-19 subject UPHS-0139

2021-03-31

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.

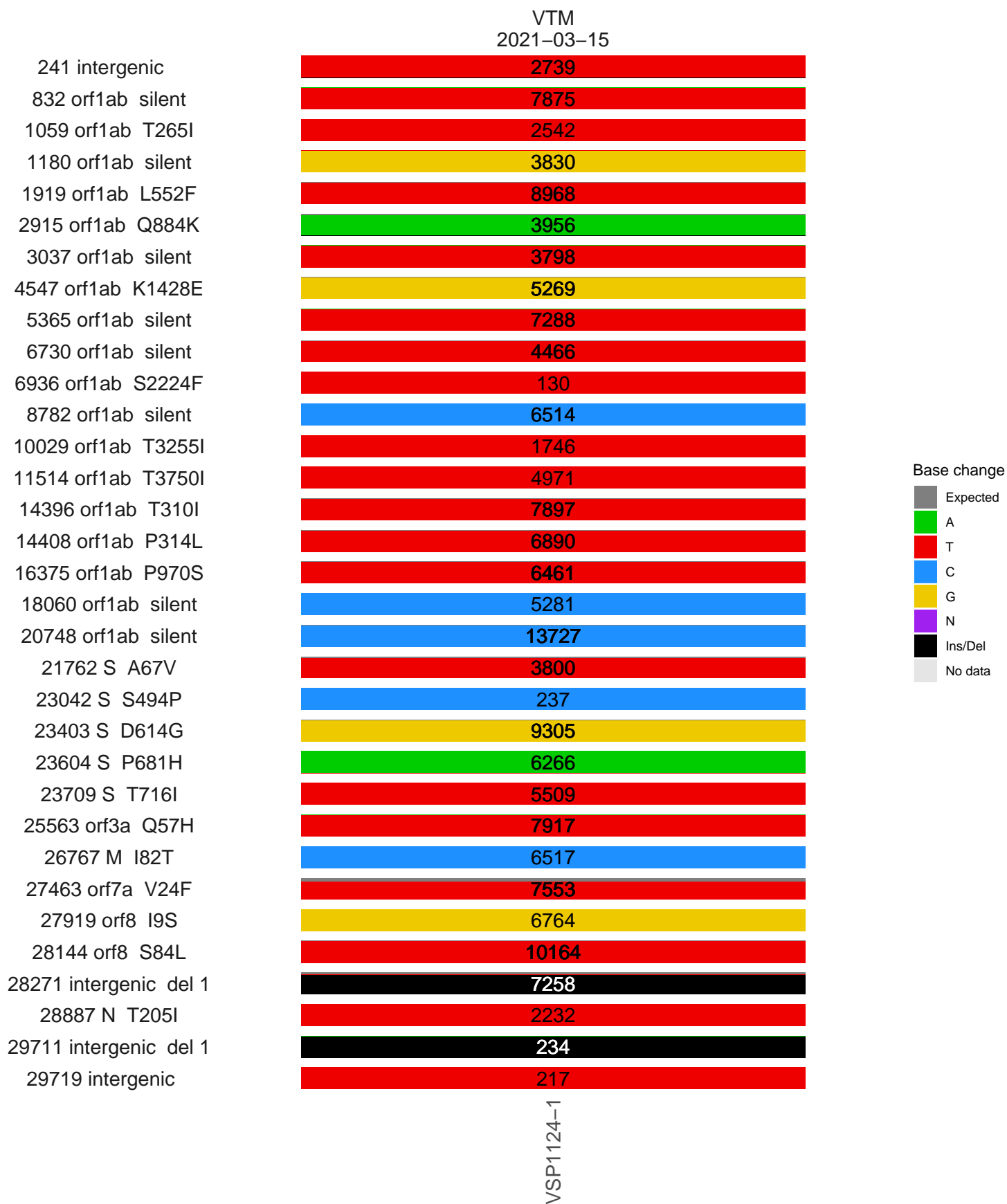Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1124-1	single experiment	NA	VTM	2021-03-15	29.83	B.1	99.8%	99.8%

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
2021-03-15

Position	VSP1124-1
241 intergenic	2739
832 orf1ab silent	7875
1059 orf1ab T265I	2542
1180 orf1ab silent	3830
1919 orf1ab L552F	8968
2915 orf1ab Q884K	3956
3037 orf1ab silent	3798
4547 orf1ab K1428E	5269
5365 orf1ab silent	7288
6730 orf1ab silent	4466
6936 orf1ab S2224F	130
8782 orf1ab silent	6514
10029 orf1ab T3255I	1746
11514 orf1ab T3750I	4971
14396 orf1ab T310I	7897
14408 orf1ab P314L	6890
16375 orf1ab P970S	6461
18060 orf1ab silent	5281
20748 orf1ab silent	13727
21762 S A67V	3800
23042 S S494P	237
23403 S D614G	9305
23604 S P681H	6266
23709 S T716I	5509
25563 orf3a Q57H	7917
26767 M I82T	6517
27463 orf7a V24F	7553
27919 orf8 I9S	6764
28144 orf8 S84L	10164
28271 intergenic del 1	7258
28887 N T205I	2232
29711 intergenic del 1	234
29719 intergenic	217

Base change

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

3

Analyses of individual experiments and composite results

VSP1124-1 | 2021-03-15 | VTM | UPHS-0139 | 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.
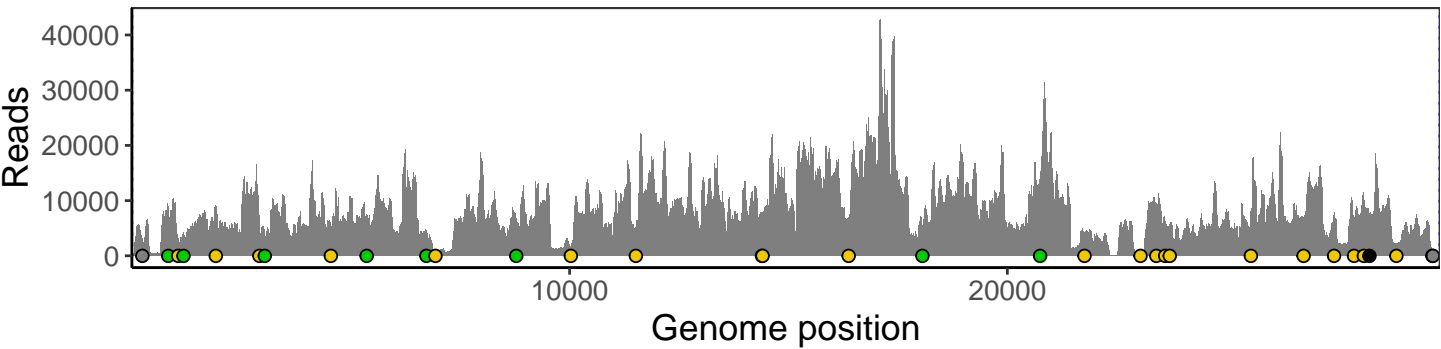
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
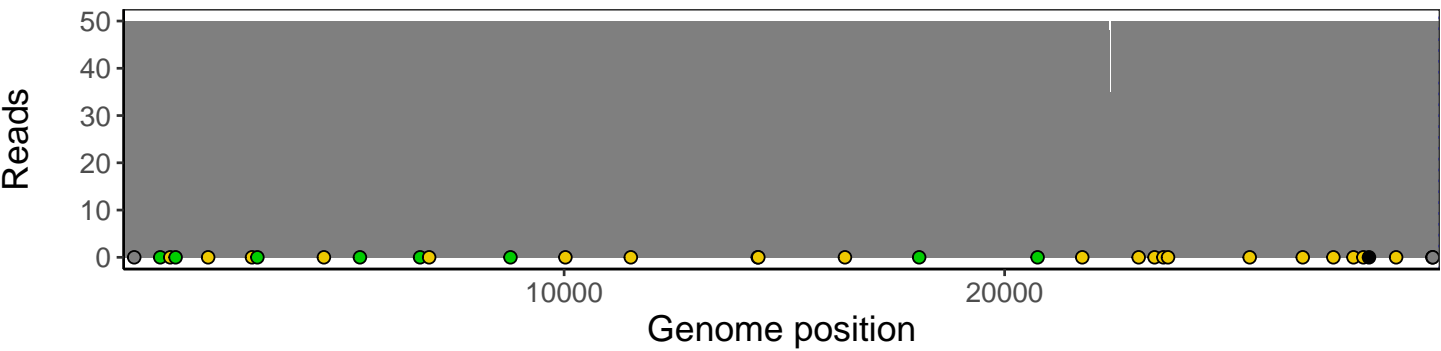
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
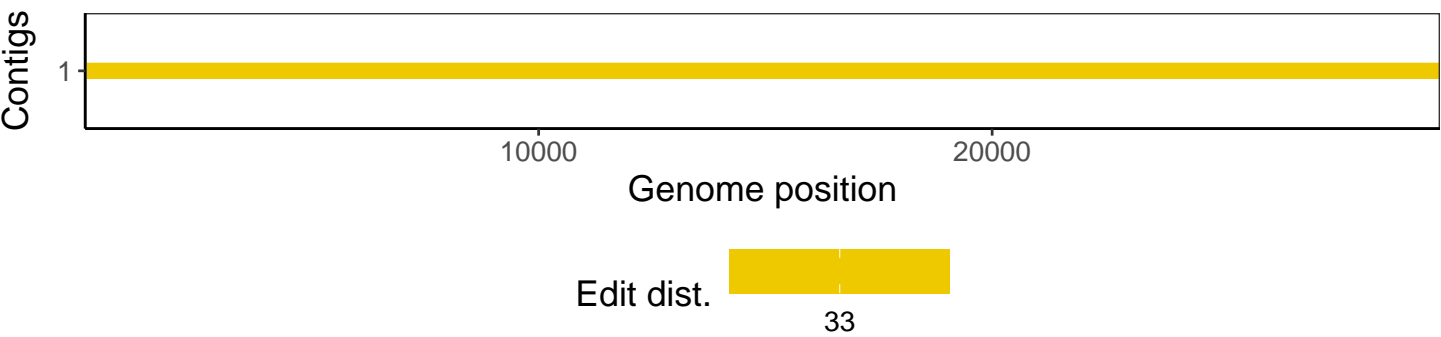
Software environment

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