COVID-19 subject UPHS-1617

2021-06-23

The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of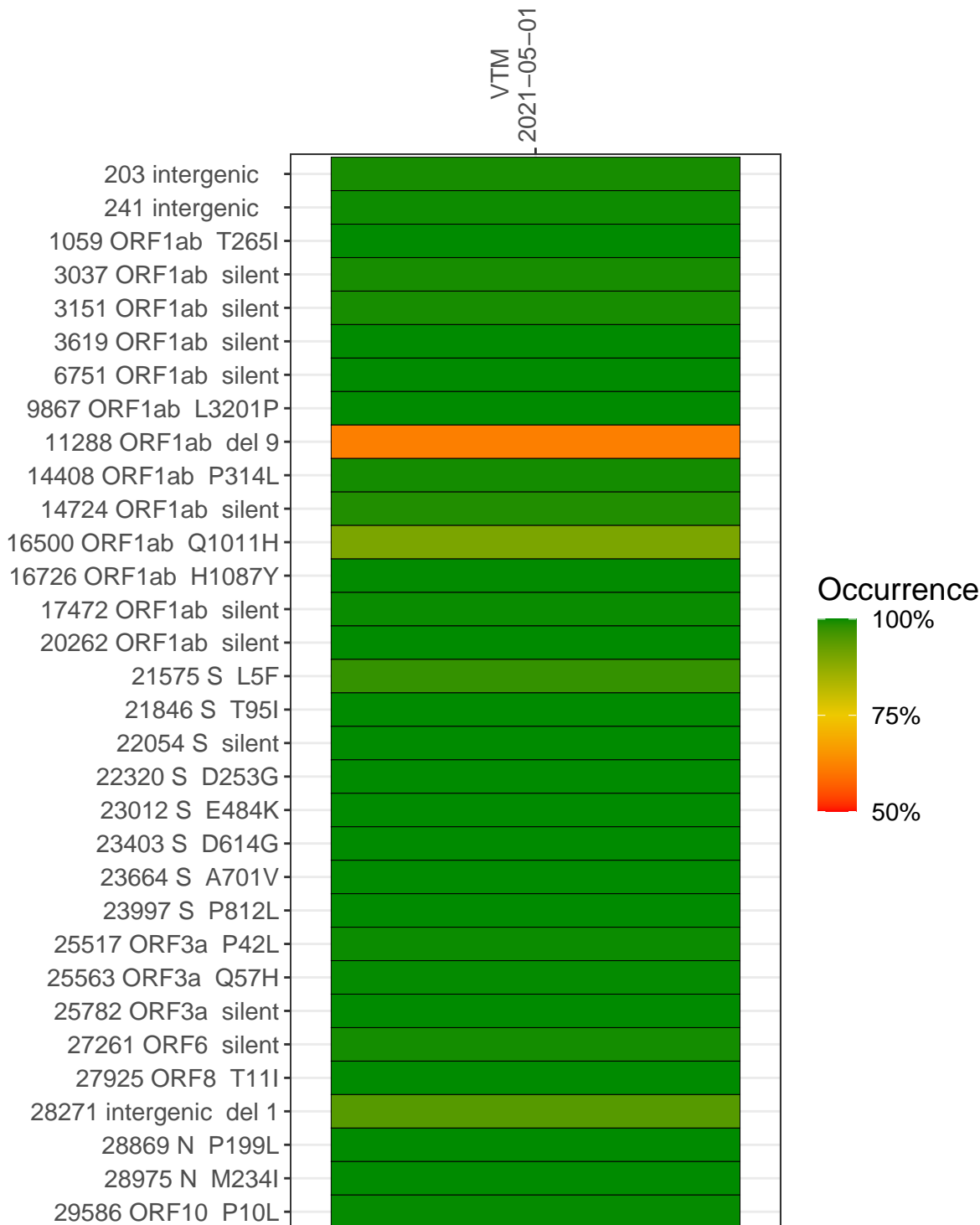 sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with > 90% sequence coverage.
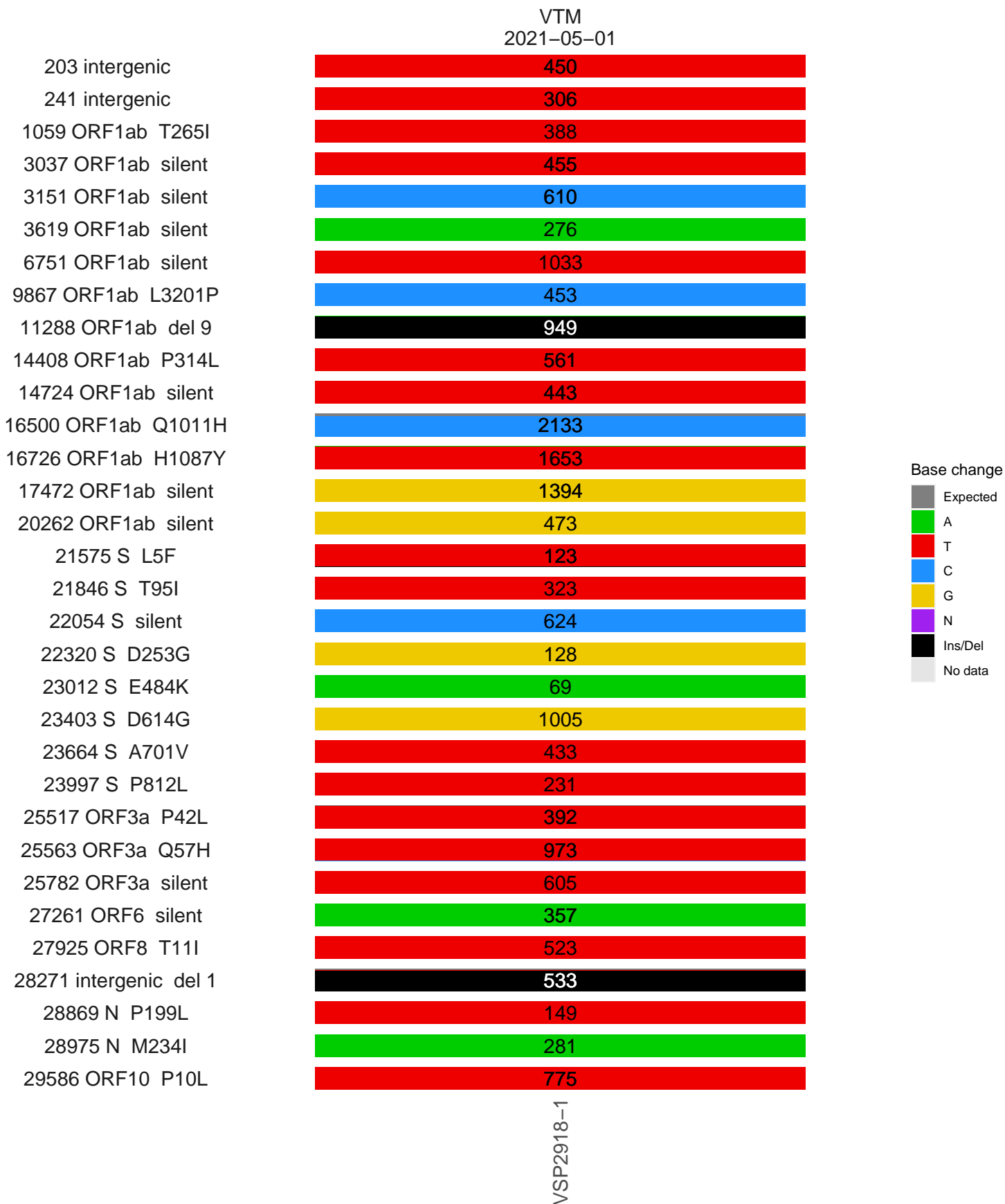
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP2918-1	single experiment	NA	VTM	2021-05-01	29.81	B.1.526	99.8%	99.5%

Variants shared across samples

The heat map below shows how variants (reference genome /home/common/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.

VTM 2021–05–01

Position	Change
203 intergenic	450 (T)
241 intergenic	306 (T)
1059 ORF1ab T265I	388 (T)
3037 ORF1ab silent	455 (T)
3151 ORF1ab silent	610 (C)
3619 ORF1ab silent	276 (A)
6751 ORF1ab silent	1033 (T)
9867 ORF1ab L3201P	453 (C)
11288 ORF1ab del 9	949 (Ins/Del)
14408 ORF1ab P314L	561 (T)
14724 ORF1ab silent	443 (T)
16500 ORF1ab Q1011H	2133 (C/Expected)
16726 ORF1ab H1087Y	1653 (T)
17472 ORF1ab silent	1394 (G)
20262 ORF1ab silent	473 (G)
21575 S L5F	123 (T)
21846 S T95I	323 (T)
22054 S silent	624 (C)
22320 S D253G	128 (G)
23012 S E484K	69 (A)
23403 S D614G	1005 (G)
23664 S A701V	433 (T)
23997 S P812L	231 (T)
25517 ORF3a P42L	392 (T)
25563 ORF3a Q57H	973 (T)
25782 ORF3a silent	605 (T)
27261 ORF6 silent	357 (A)
27925 ORF8 T11I	523 (T)
28271 intergenic del 1	533 (Ins/Del)
28869 N P199L	149 (T)
28975 N M234I	281 (A)
29586 ORF10 P10L	775 (T)

VSP2918–1

Base change

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

3

Analyses of individual experiments and composite results

**VSP2918-1 | 2021-05-01 | VTM | UPHS-1617 | 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.
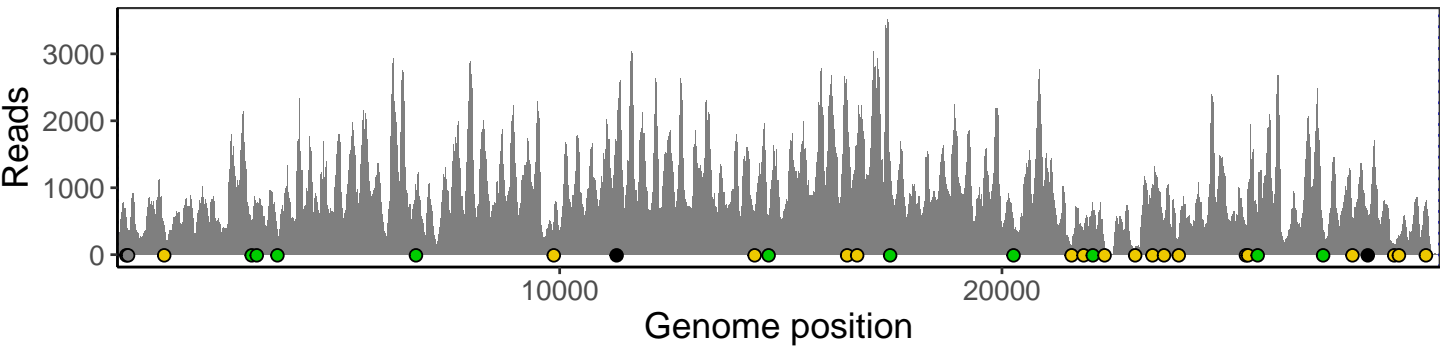
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
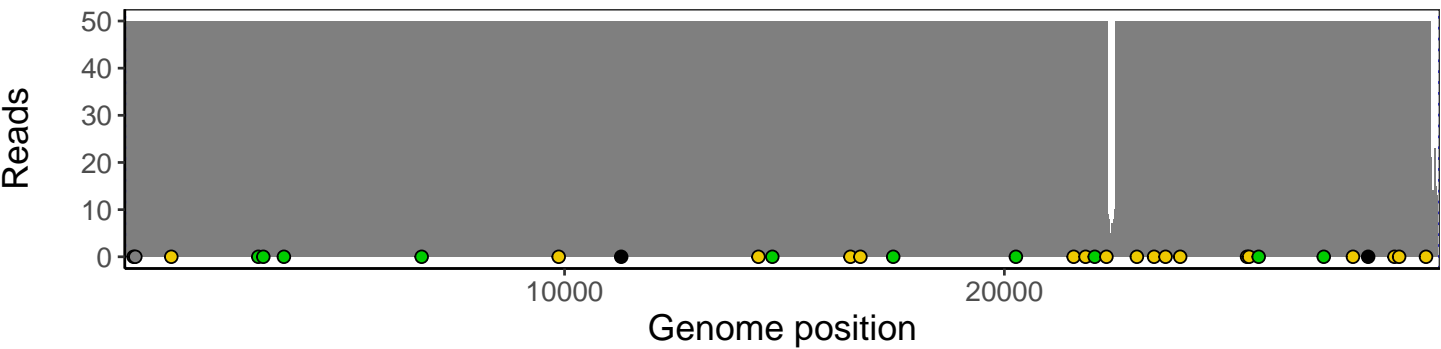
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
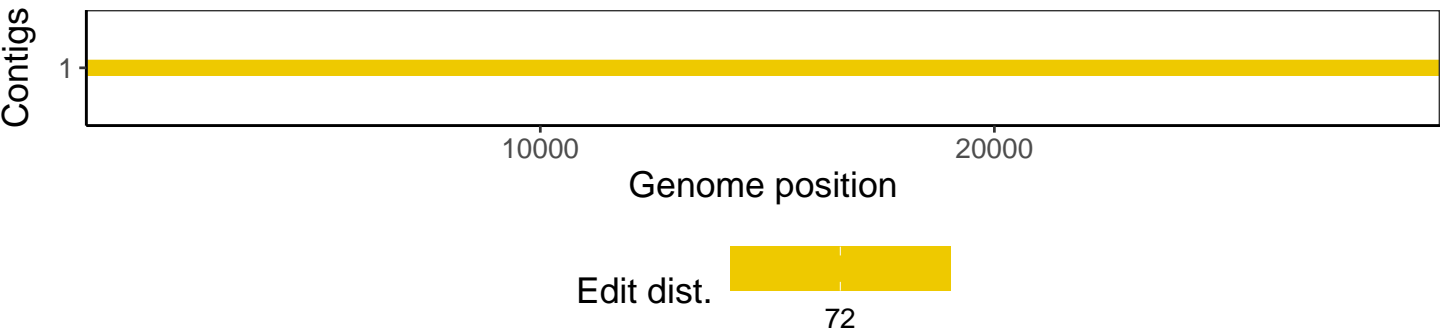
Software environment

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	3.1.3
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	3.3.3
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