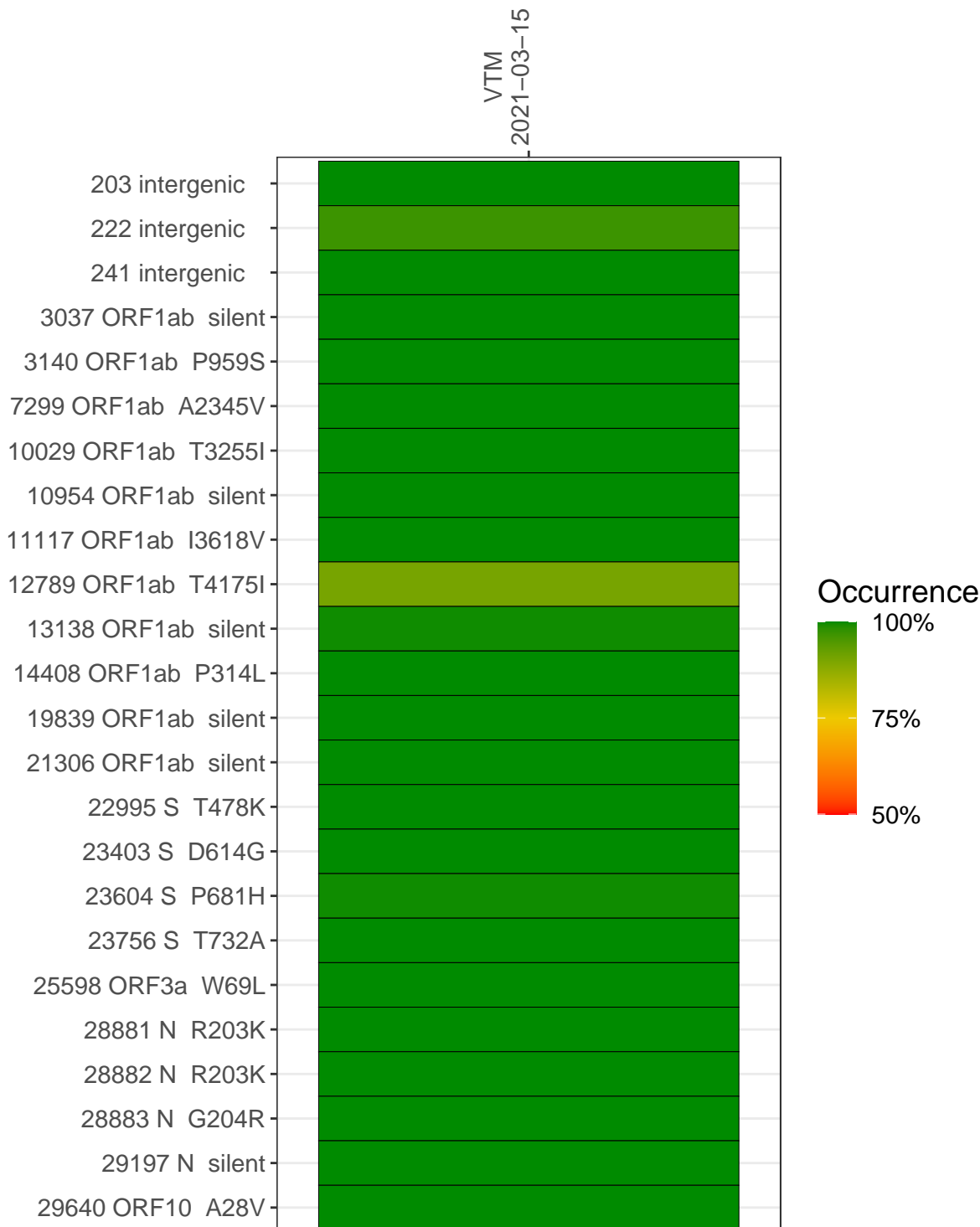# COVID-19 subject UPHS-0192

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.
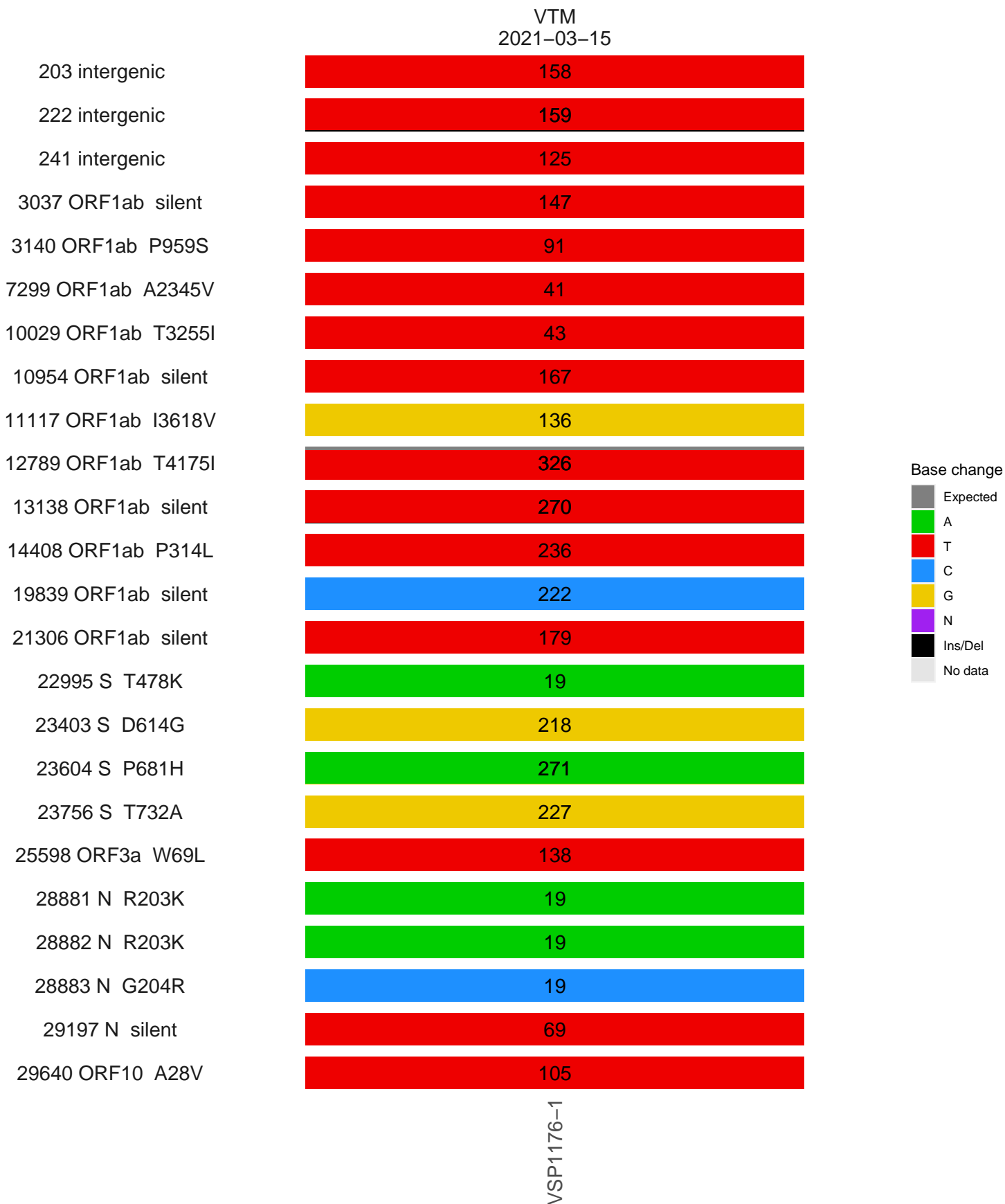
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1176-1	single experiment	NA	VTM	2021-03-15	22.36	B.1.1.519	99.8%	98.9%

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.

VTM
2021−03−15

Position / Gene / Mutation	Value	Base
203 intergenic	158	T
222 intergenic	159	T
241 intergenic	125	T
3037 ORF1ab silent	147	T
3140 ORF1ab P959S	91	T
7299 ORF1ab A2345V	41	T
10029 ORF1ab T3255I	43	T
10954 ORF1ab silent	167	T
11117 ORF1ab I3618V	136	G
12789 ORF1ab T4175I	326	T
13138 ORF1ab silent	270	T
14408 ORF1ab P314L	236	T
19839 ORF1ab silent	222	C
21306 ORF1ab silent	179	T
22995 S T478K	19	A
23403 S D614G	218	G
23604 S P681H	271	A
23756 S T732A	227	G
25598 ORF3a W69L	138	T
28881 N R203K	19	A
28882 N R203K	19	A
28883 N G204R	19	C
29197 N silent	69	T
29640 ORF10 A28V	105	T

VSP1176−1

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

Analyses of individual experiments and composite results

## VSP1176-1 | 2021-03-15 | VTM | UPHS-0192 | 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.
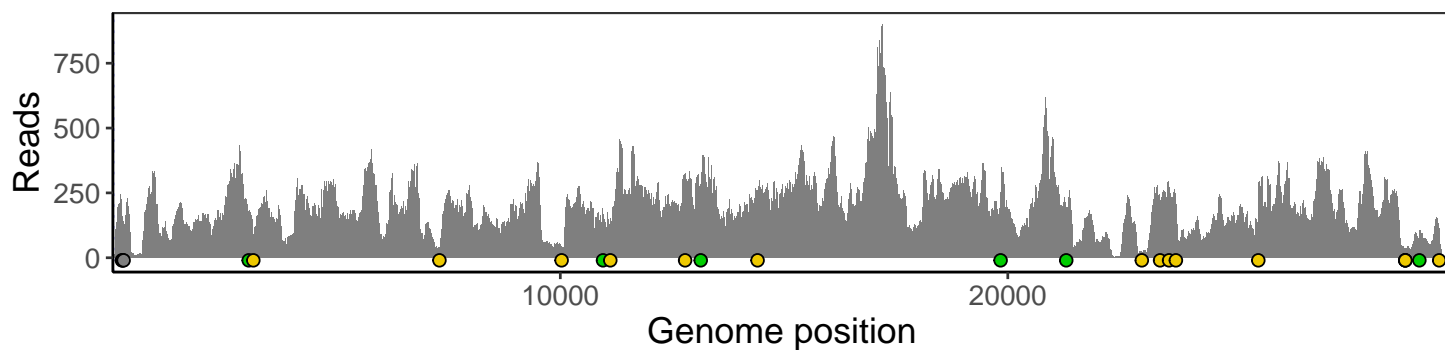
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
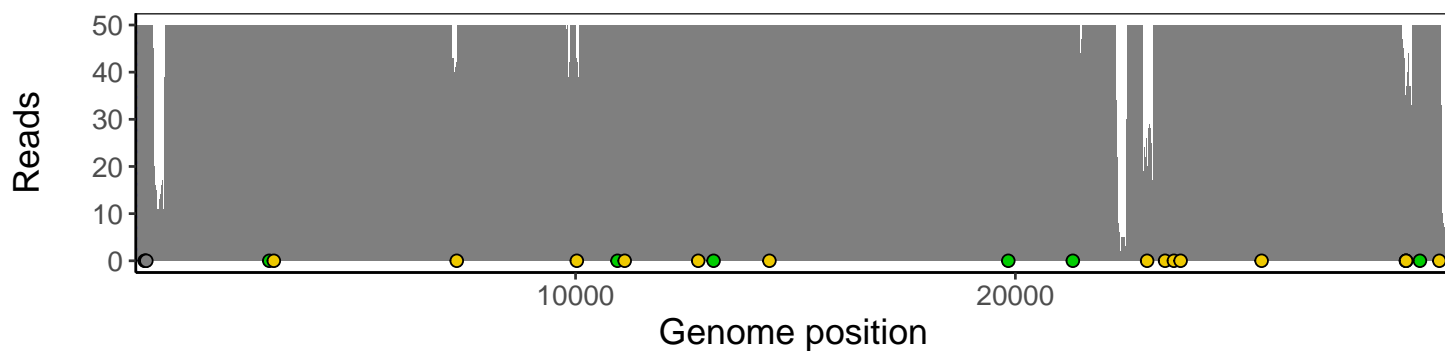
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
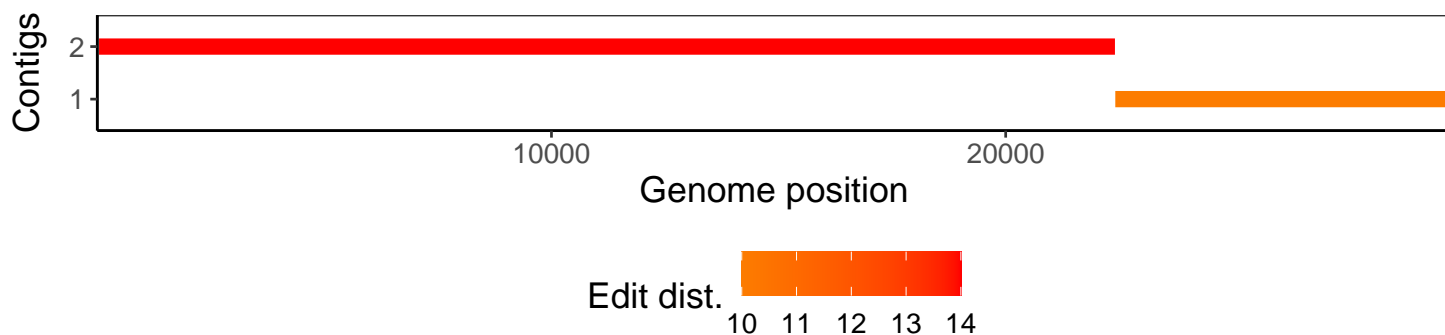
Software environment

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	3.0.0
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