COVID-19 subject UPHS-0812

2021-06-23

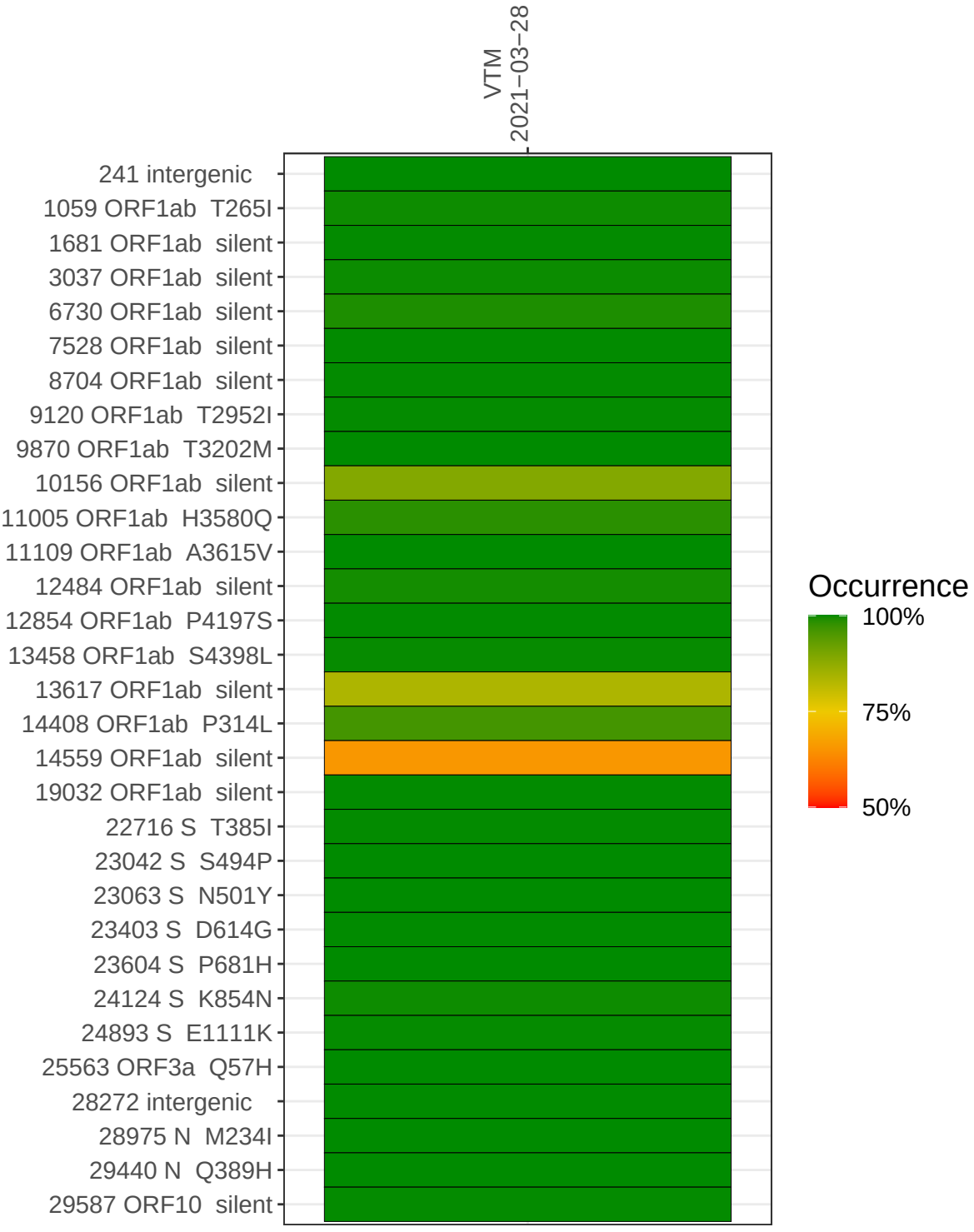
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

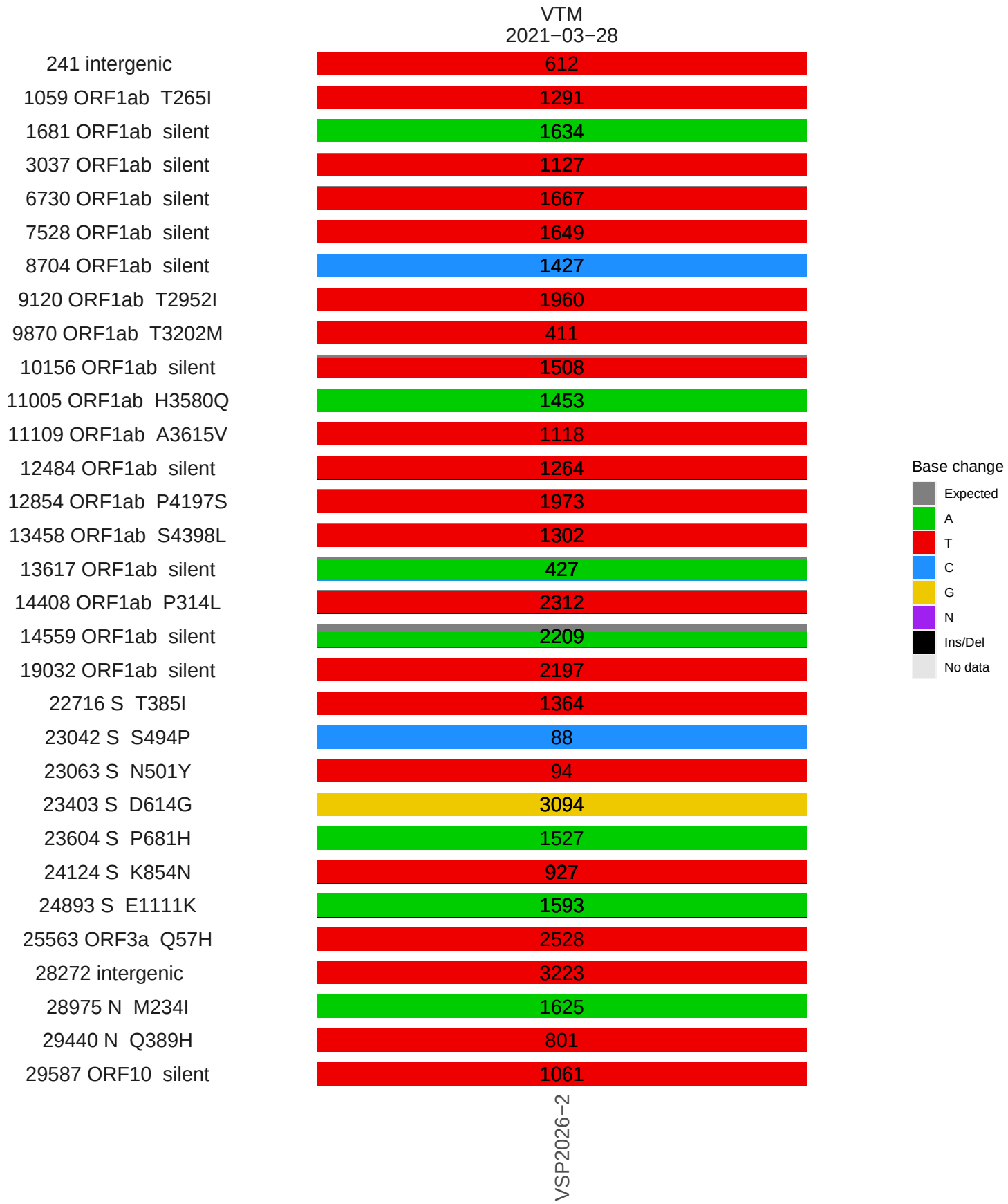
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP2026-2	single experiment	NA	VTM	2021-03-28	29.65	B.1.623	99.1%	99.1%

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.

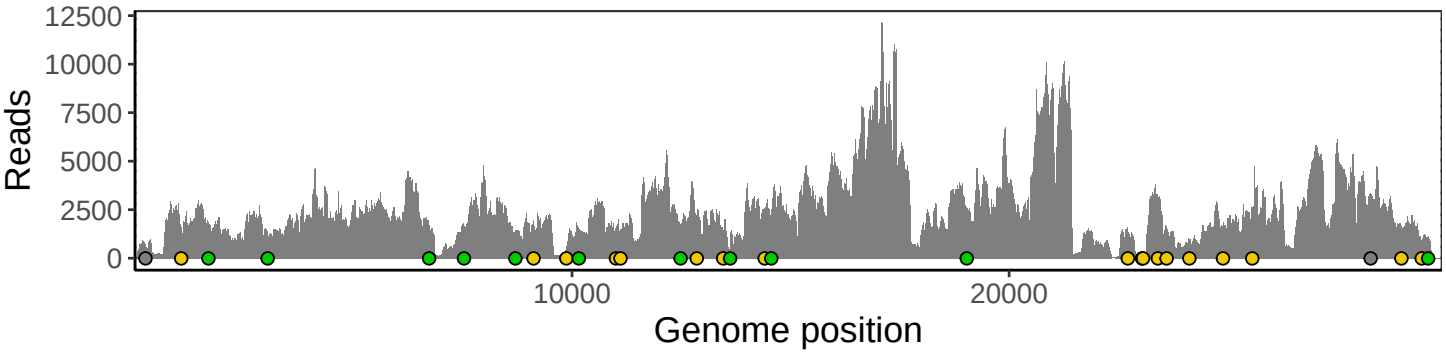



VTM
2021−03−28
241 intergenic
612
1059 ORF1ab T265I
1291
1681 ORF1ab silent
1634
3037 ORF1ab silent
1127
6730 ORF1ab silent
1667
7528 ORF1ab silent
1649
8704 ORF1ab silent
1427
9120 ORF1ab T2952I
1960
9870 ORF1ab T3202M
411
10156 ORF1ab silent
1508
11005 ORF1ab H3580Q
1453
11109 ORF1ab A3615V
1118
12484 ORF1ab silent
1264
12854 ORF1ab P4197S
1973
13458 ORF1ab S4398L
1302
13617 ORF1ab silent
427
14408 ORF1ab P314L
2312
14559 ORF1ab silent
2209
19032 ORF1ab silent
2197
22716 S T385I
1364
23042 S S494P
88
23063 S N501Y
94
23403 S D614G
3094
23604 S P681H
1527
24124 S K854N
927
24893 S E1111K
1593
25563 ORF3a Q57H
2528
28272 intergenic
3223
28975 N M234I
1625
29440 N Q389H
801
29587 ORF10 silent
1061
VSP2026−2
Base change
Expected
A
T
C
G
N
Ins/Del
No data

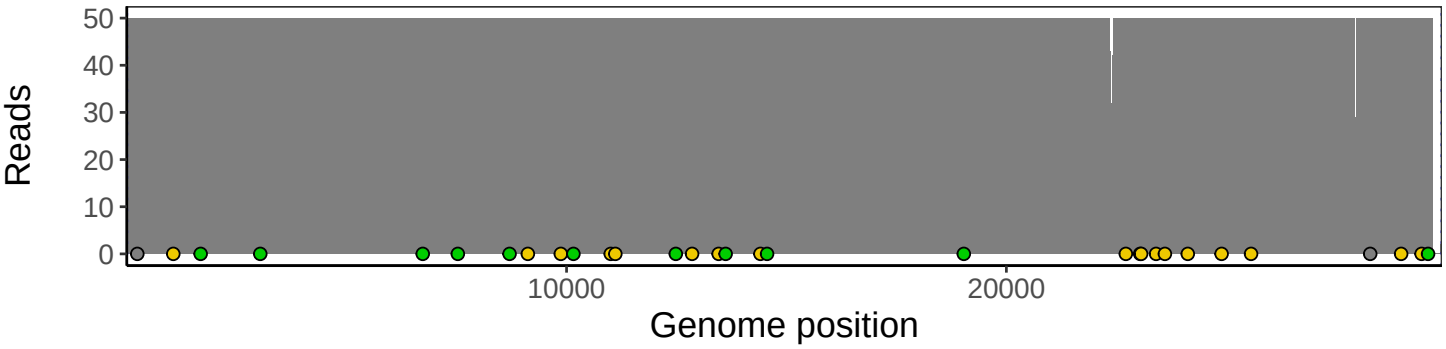
Analyses of individual experiments and composite results

VSP2026-2 | 2021-03-28 | VTM | UPHS-0812 | genomes | single experiment

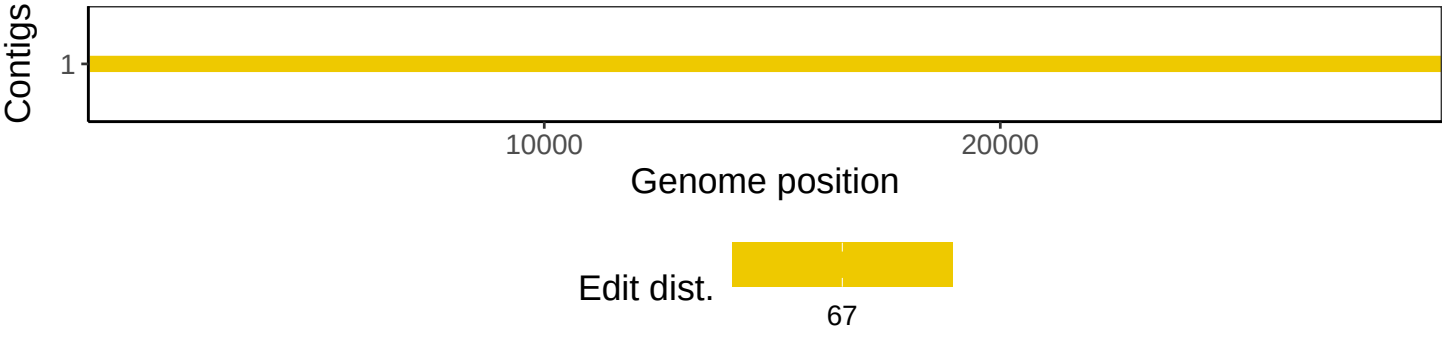
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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