COVID-19 subject UPHS-0262

2021-04-17

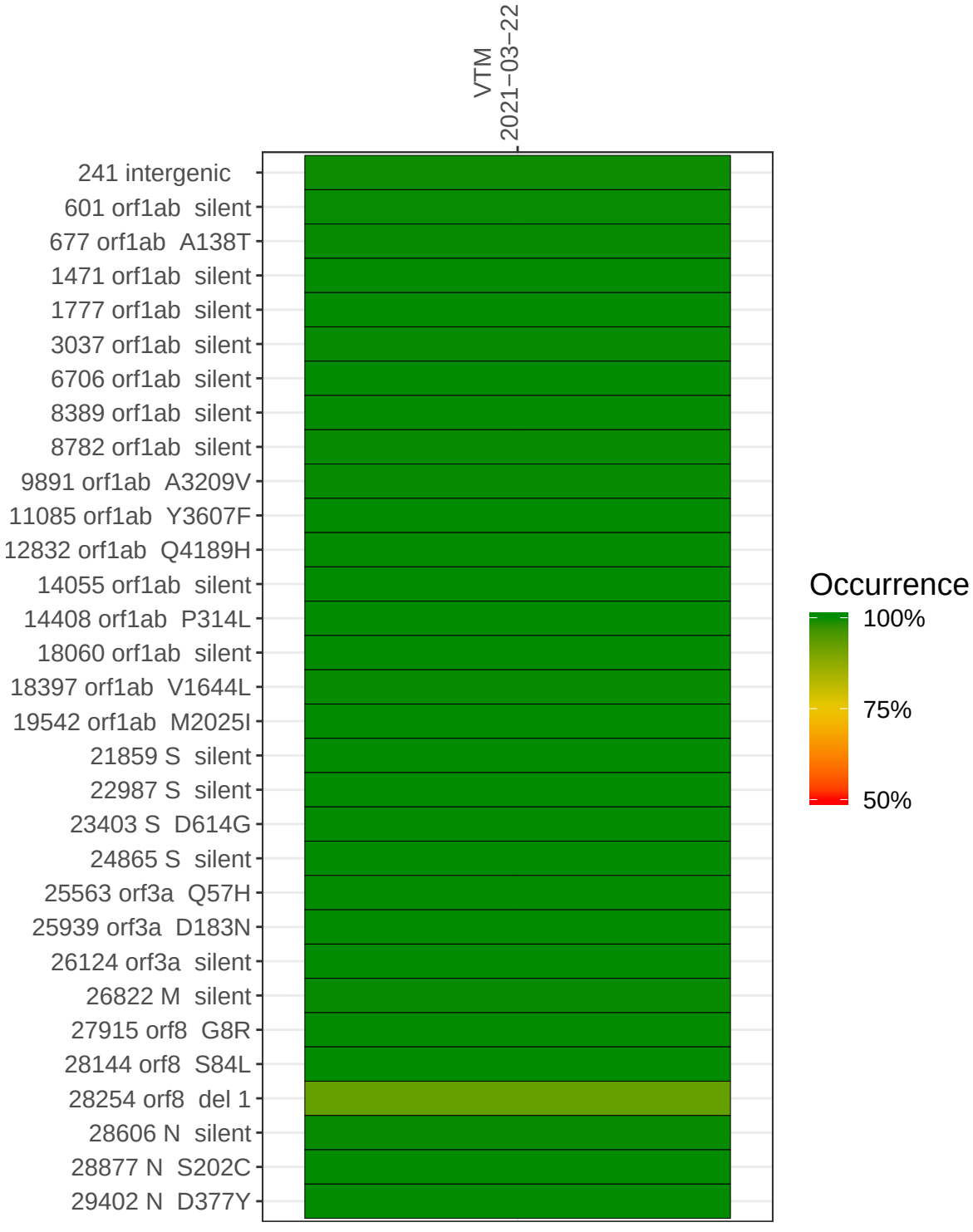
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

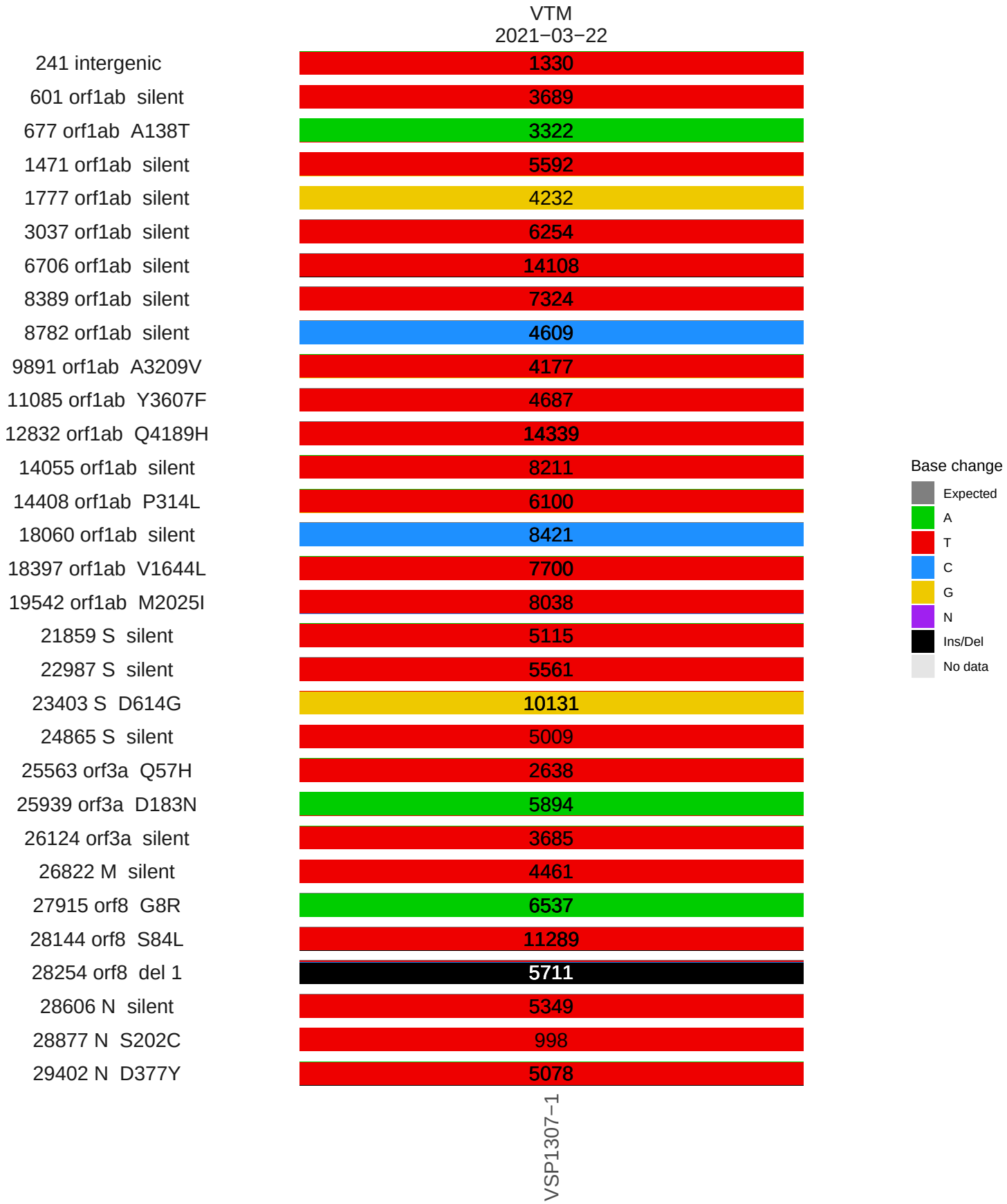
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1307-1	single experiment	NA	VTM	2021-03-22	29.88	B.1.110.3	99.9%	99.8%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/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.

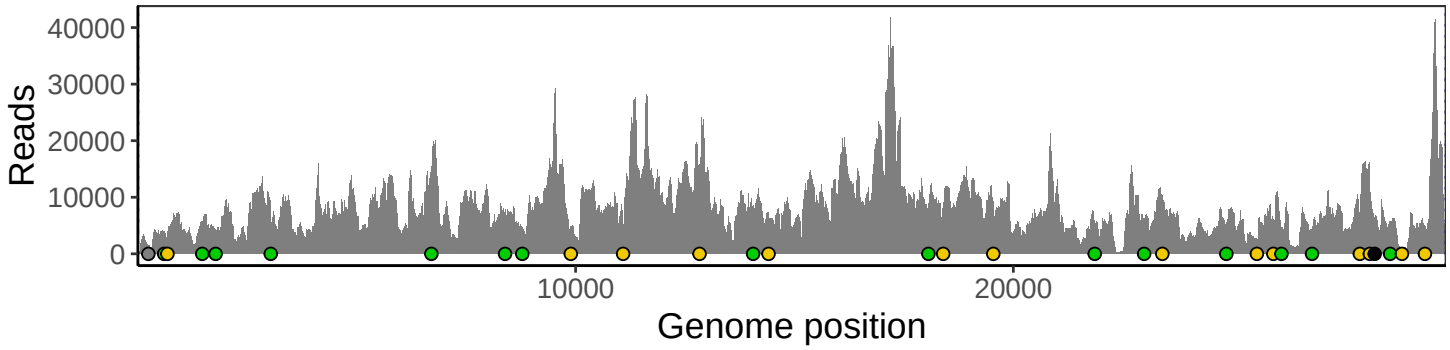



VTM
2021−03−22
241 intergenic
601 orf1ab silent
677 orf1ab A138T
1471 orf1ab silent
1777 orf1ab silent
3037 orf1ab silent
6706 orf1ab silent
8389 orf1ab silent
8782 orf1ab silent
9891 orf1ab A3209V
11085 orf1ab Y3607F
12832 orf1ab Q4189H
14055 orf1ab silent
14408 orf1ab P314L
18060 orf1ab silent
18397 orf1ab V1644L
19542 orf1ab M2025I
21859 S silent
22987 S silent
23403 S D614G
24865 S silent
25563 orf3a Q57H
25939 orf3a D183N
26124 orf3a silent
26822 M silent
27915 orf8 G8R
28144 orf8 S84L
28254 orf8 del 1
28606 N silent
28877 N S202C
29402 N D377Y
1330
3689
3322
5592
4232
6254
14108
7324
4609
4177
4687
14339
8211
6100
8421
7700
8038
5115
5561
10131
5009
2638
5894
3685
4461
6537
11289
5711
5349
998
5078
VSP1307−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

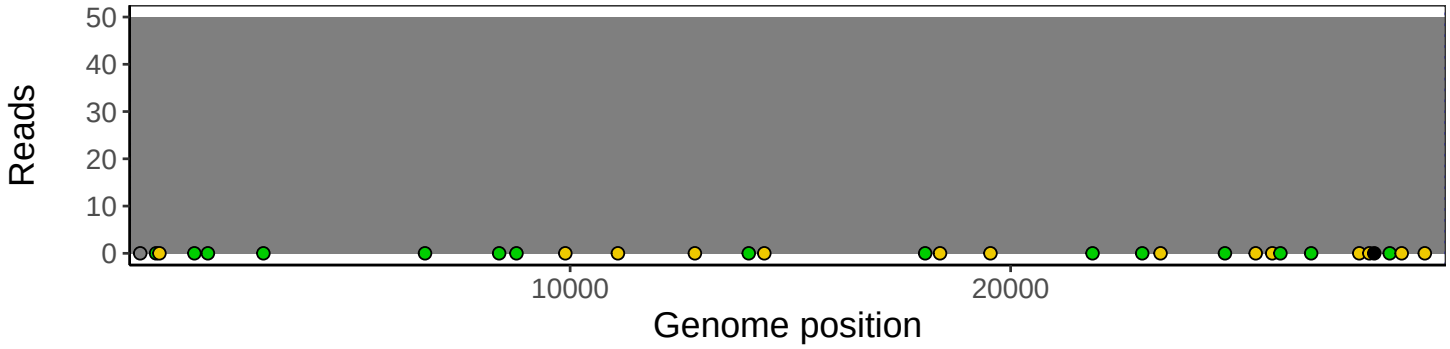
Analyses of individual experiments and composite results

VSP1307-1 | 2021-03-22 | VTM | UPHS-0262 | genomes | single experiment

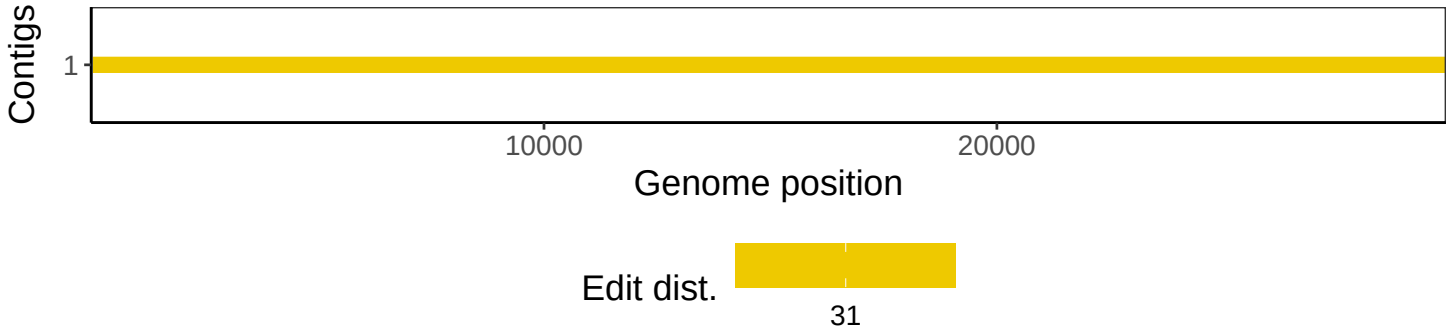
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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