COVID-19 subject UPHS-1159

2021-06-23

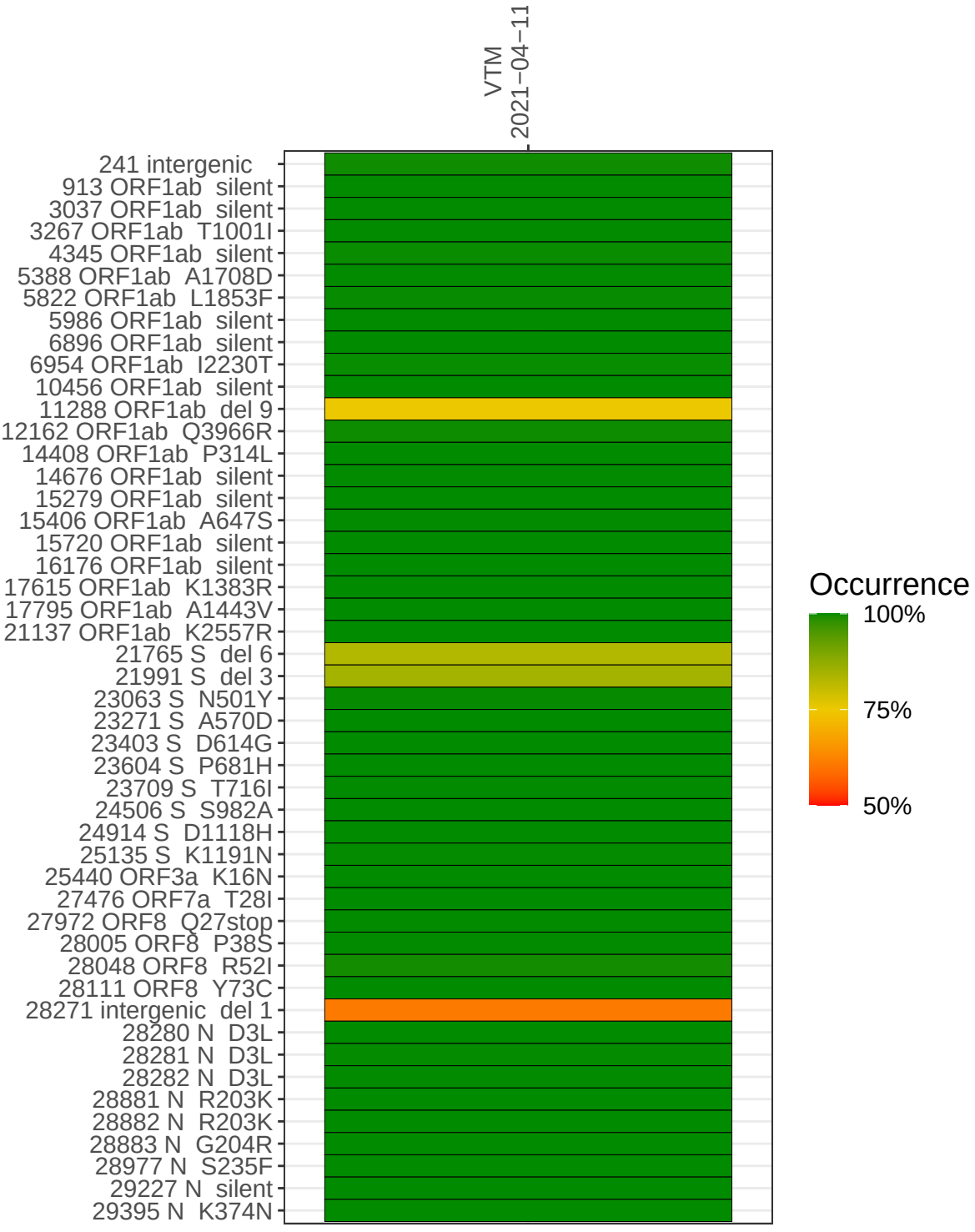
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

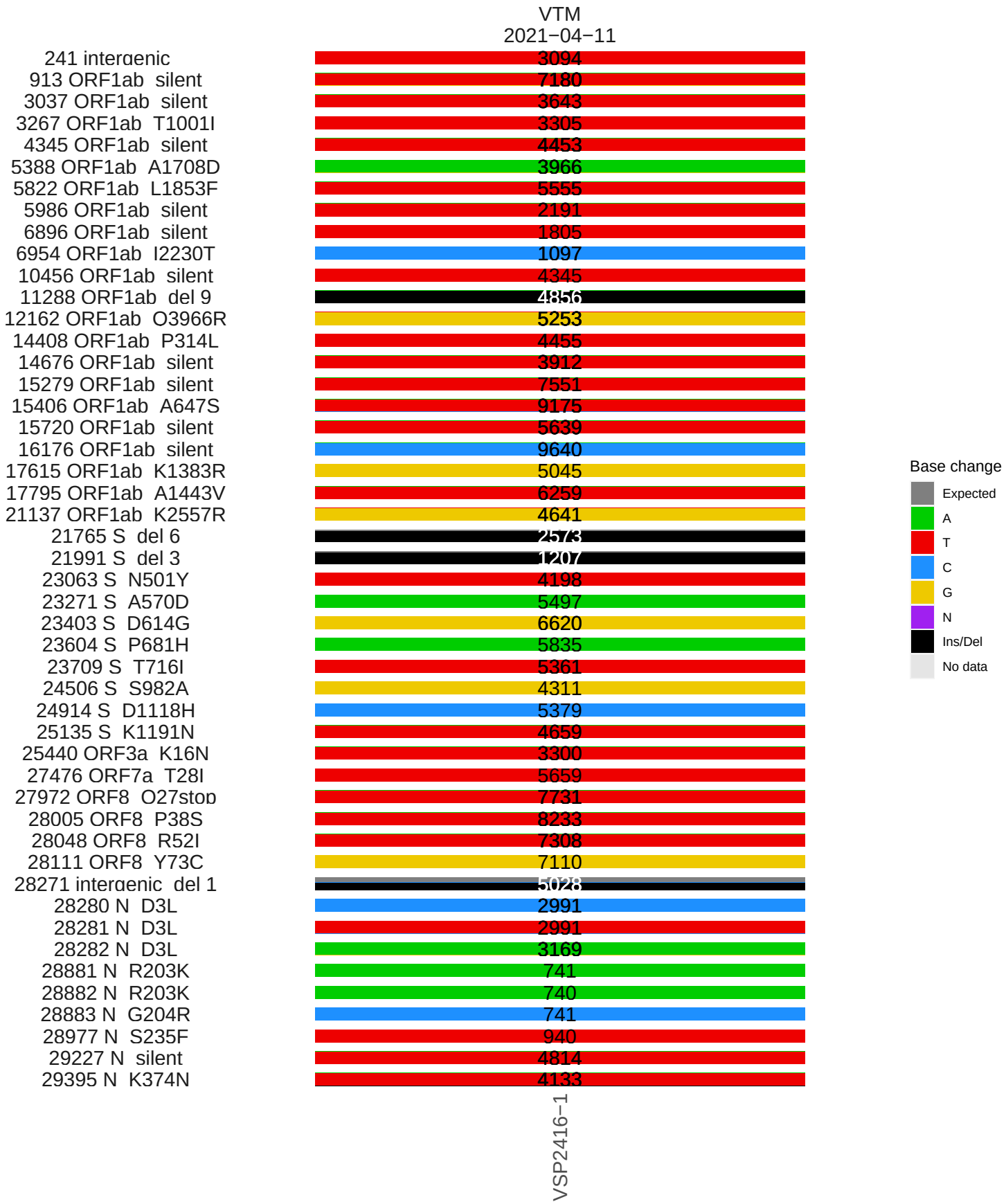
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP2416-1	single experiment	NA	VTM	2021-04-11	29.81	B.1.1.7	99.8%	99.7%

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.

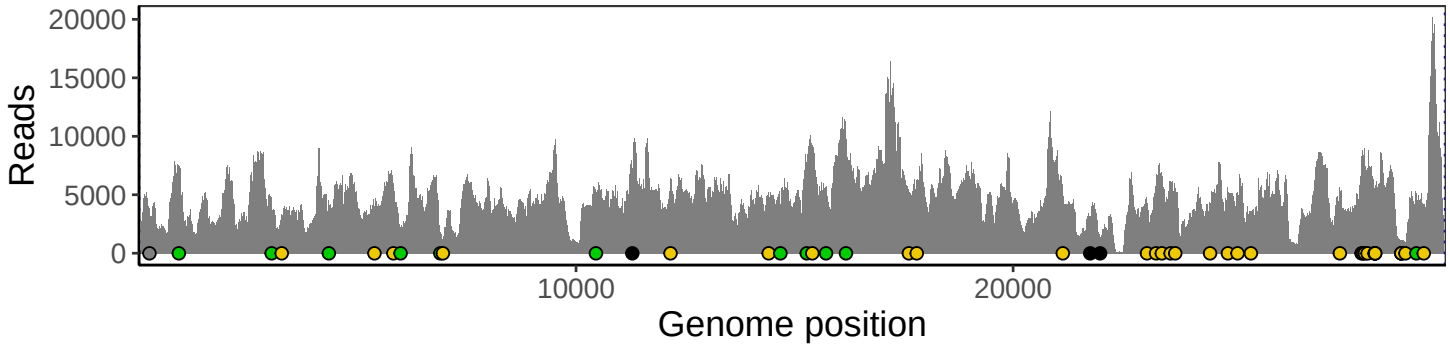



VTM
2021−04−11
241 intergenic 3094
913 ORF1ab silent 7180
3037 ORF1ab silent 3643
3267 ORF1ab T1001I 3305
4345 ORF1ab silent 4453
5388 ORF1ab A1708D 3966
5822 ORF1ab L1853F 5555
5986 ORF1ab silent 2191
6896 ORF1ab silent 1805
6954 ORF1ab I2230T 1097
10456 ORF1ab silent 4345
11288 ORF1ab del 9 4856
12162 ORF1ab O3966R 5253
14408 ORF1ab P314L 4455
14676 ORF1ab silent 3912
15279 ORF1ab silent 7551
15406 ORF1ab A647S 9175
15720 ORF1ab silent 5639
16176 ORF1ab silent 9640
17615 ORF1ab K1383R 5045
17795 ORF1ab A1443V 6259
21137 ORF1ab K2557R 4641
21765 S del 6 2573
21991 S del 3 1207
23063 S N501Y 4198
23271 S A570D 5497
23403 S D614G 6620
23604 S P681H 5835
23709 S T716I 5361
24506 S S982A 4311
24914 S D1118H 5379
25135 S K1191N 4659
25440 ORF3a K16N 3300
27476 ORF7a T28I 5659
27972 ORF8 O27stop 7731
28005 ORF8 P38S 8233
28048 ORF8 R52I 7308
28111 ORF8 Y73C 7110
28271 intergenic del 1 5028
28280 N D3L 2991
28281 N D3L 2991
28282 N D3L 3169
28881 N R203K 741
28882 N R203K 740
28883 N G204R 741
28977 N S235F 940
29227 N silent 4814
29395 N K374N 4133
VSP2416−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

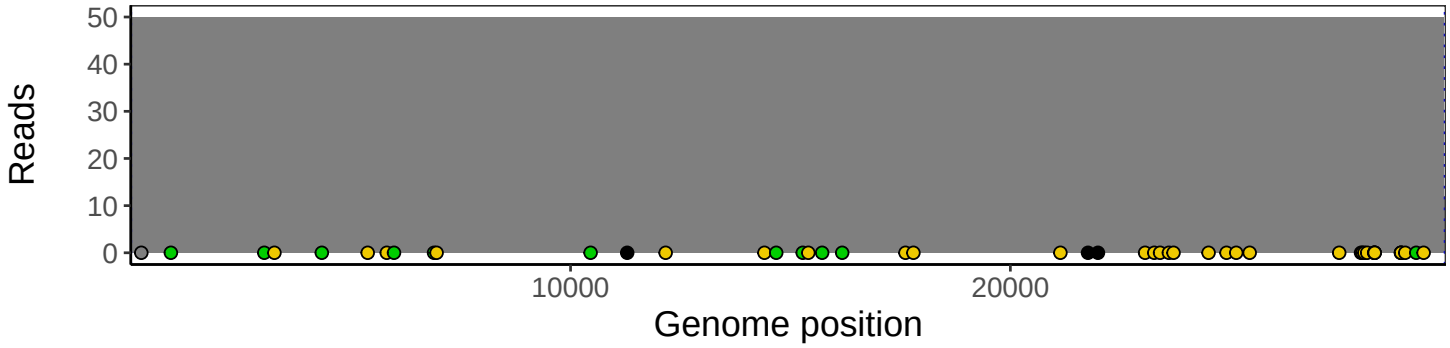
Analyses of individual experiments and composite results

VSP2416-1 | 2021-04-11 | VTM | UPHS-1159 | genomes | single experiment

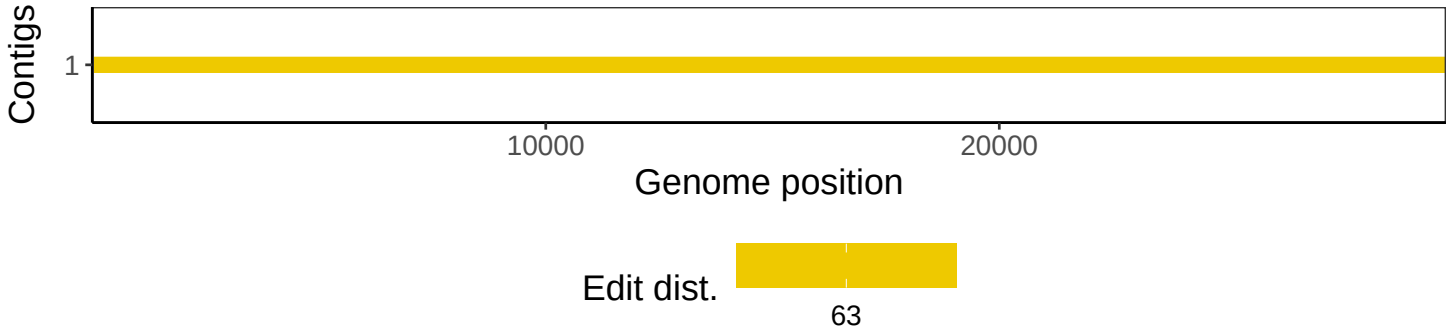
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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