COVID-19 subject UPHS-1654

2021-06-03

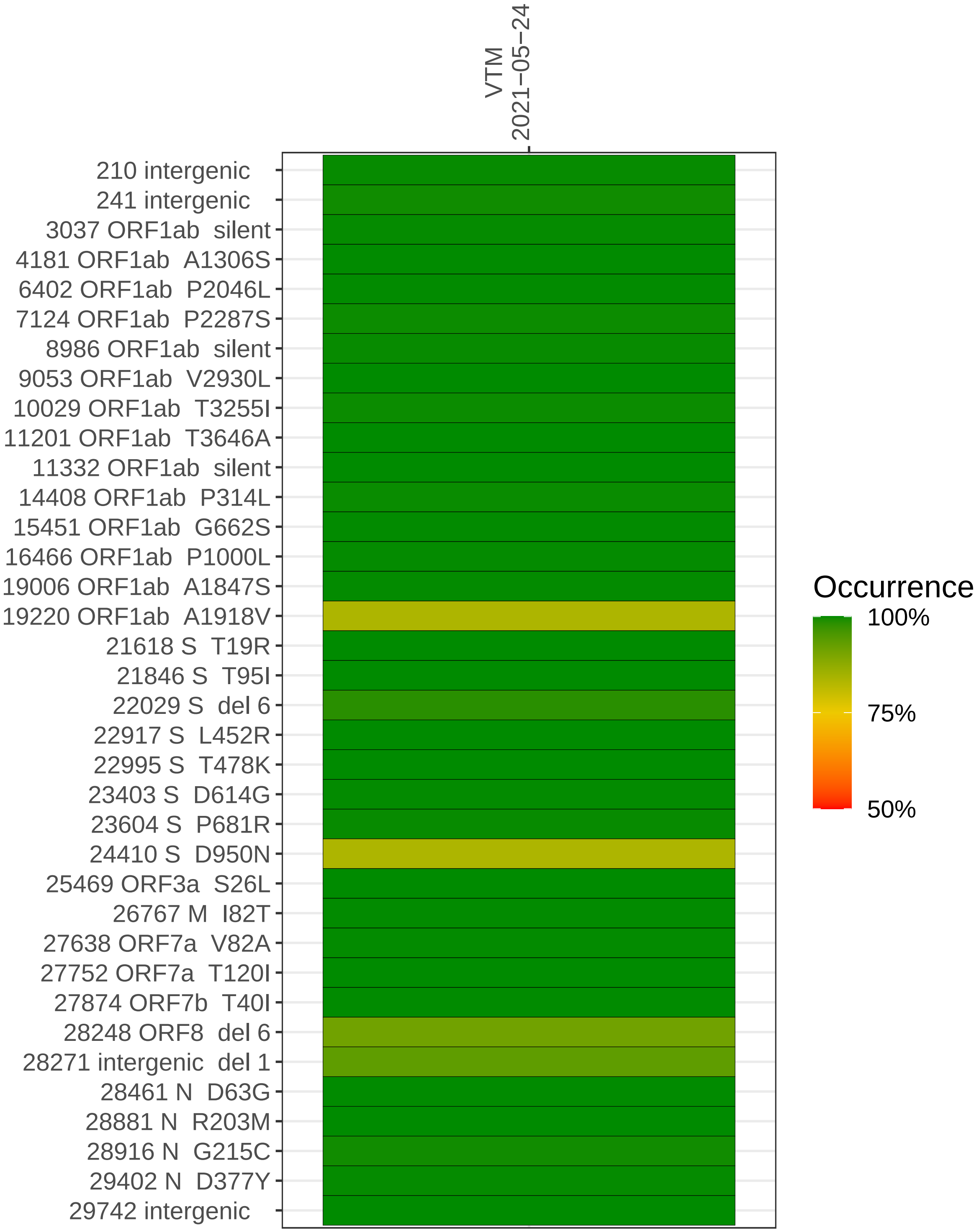
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

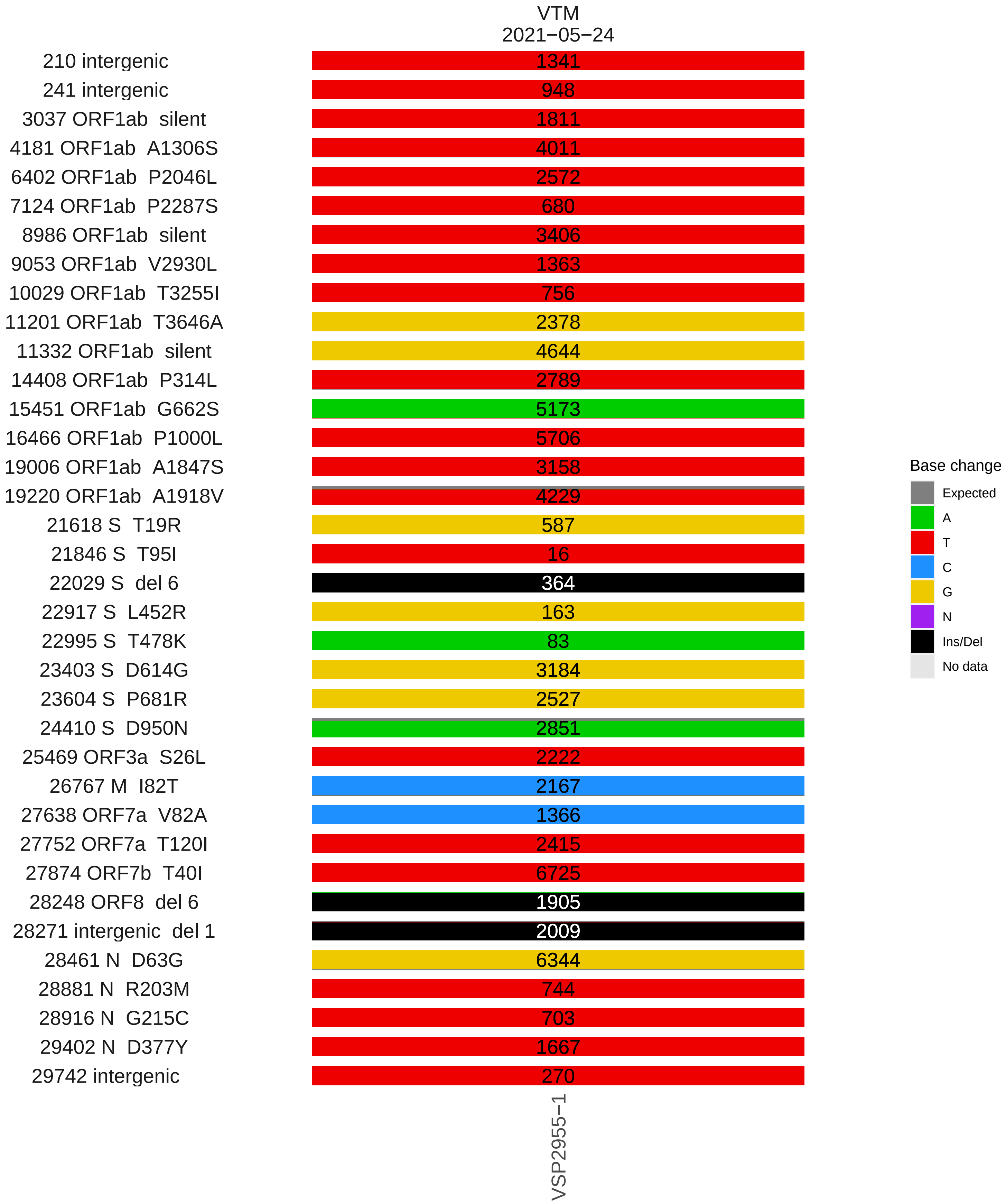
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP2955-1	single experiment	NA	VTM	2021-05-24	29.82	B.1.617	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.

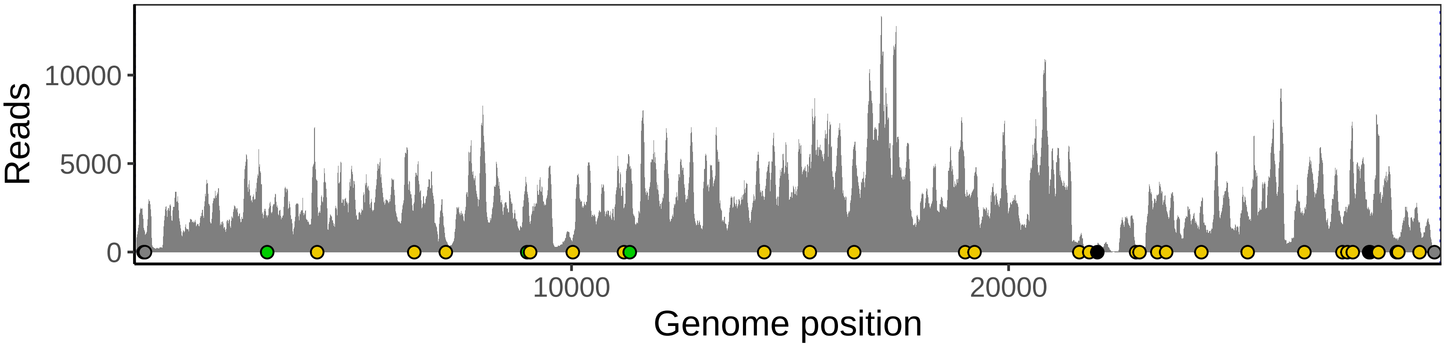



VTM
2021−05−24
210 intergenic
1341
241 intergenic
948
3037 ORF1ab silent
1811
4181 ORF1ab A1306S
4011
6402 ORF1ab P2046L
2572
7124 ORF1ab P2287S
680
8986 ORF1ab silent
3406
9053 ORF1ab V2930L
1363
10029 ORF1ab T3255I
756
11201 ORF1ab T3646A
2378
11332 ORF1ab silent
4644
14408 ORF1ab P314L
2789
15451 ORF1ab G662S
5173
16466 ORF1ab P1000L
5706
19006 ORF1ab A1847S
3158
19220 ORF1ab A1918V
4229
21618 S T19R
587
21846 S T95I
16
22029 S del 6
364
22917 S L452R
163
22995 S T478K
83
23403 S D614G
3184
23604 S P681R
2527
24410 S D950N
2851
25469 ORF3a S26L
2222
26767 M I82T
2167
27638 ORF7a V82A
1366
27752 ORF7a T120I
2415
27874 ORF7b T40I
6725
28248 ORF8 del 6
1905
28271 intergenic del 1
2009
28461 N D63G
6344
28881 N R203M
744
28916 N G215C
703
29402 N D377Y
1667
29742 intergenic
270
VSP2955−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

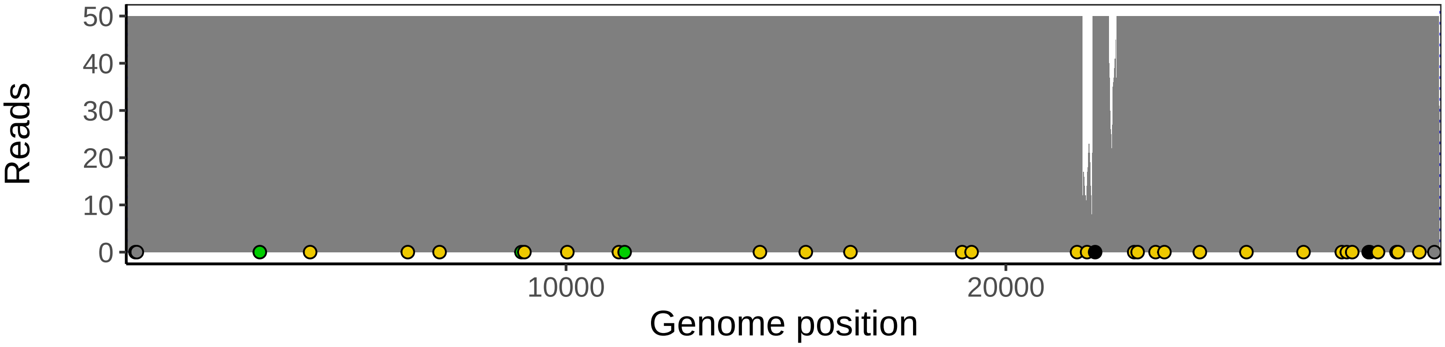
Analyses of individual experiments and composite results

VSP2955-1 | 2021-05-24 | VTM | UPHS-1654 | genomes | single experiment

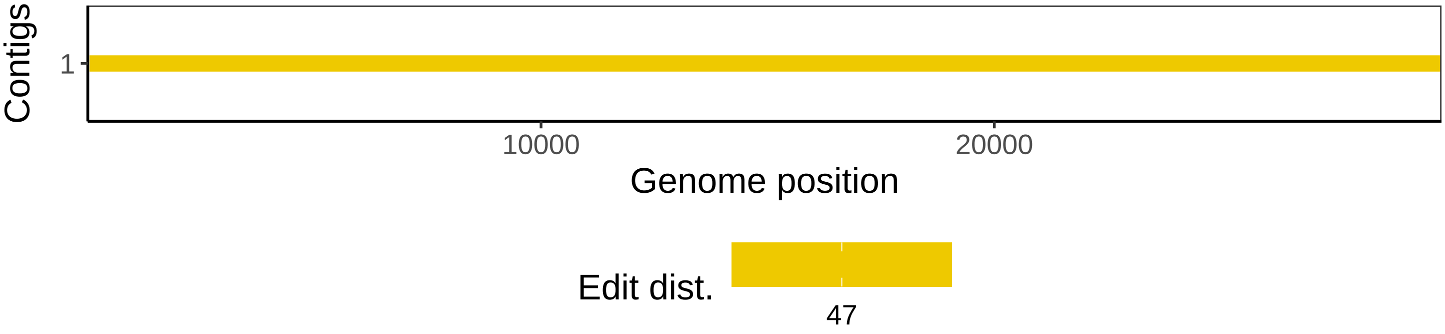
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	3.3.3
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