COVID-19 subject H2103110830

2021-05-05

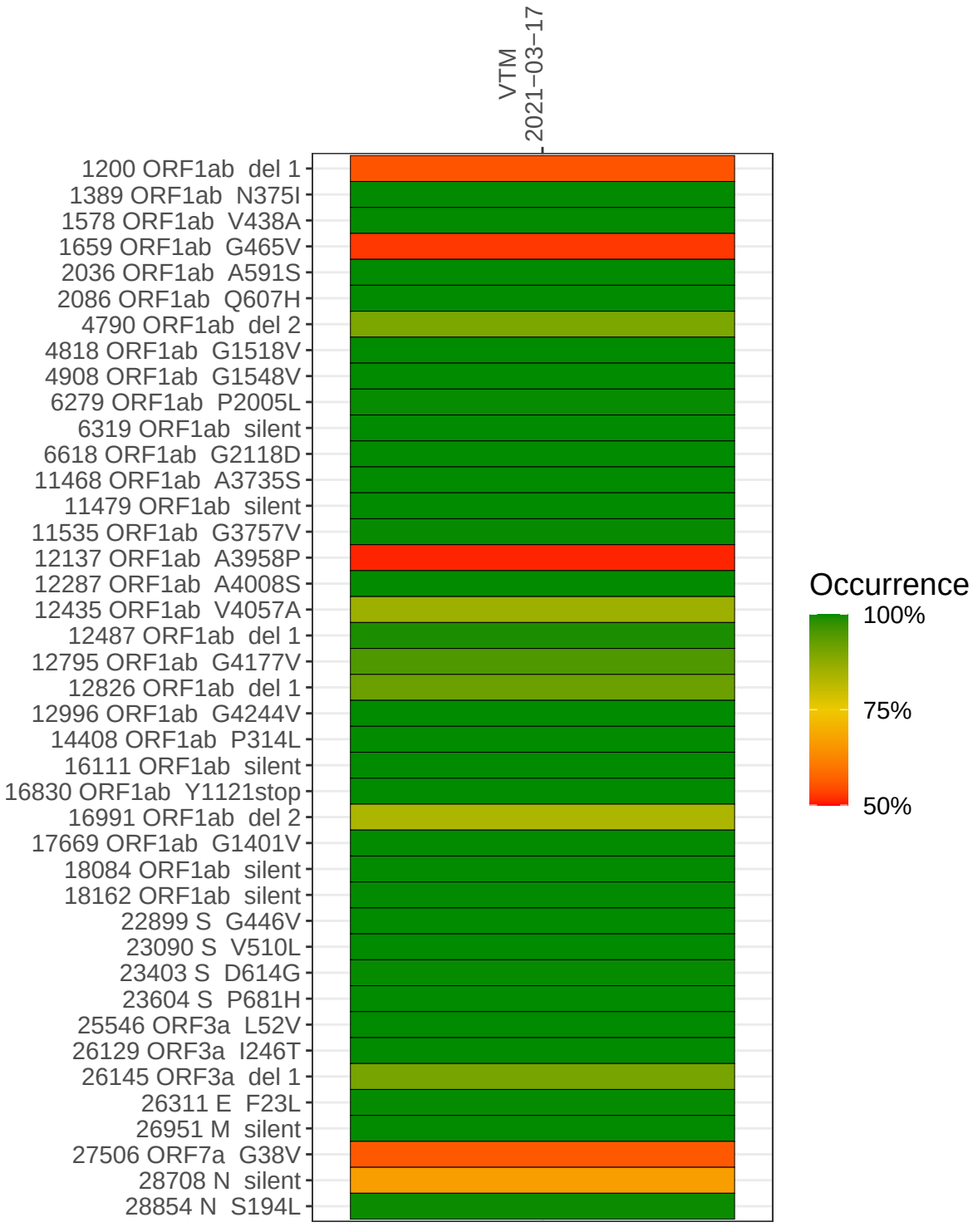
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

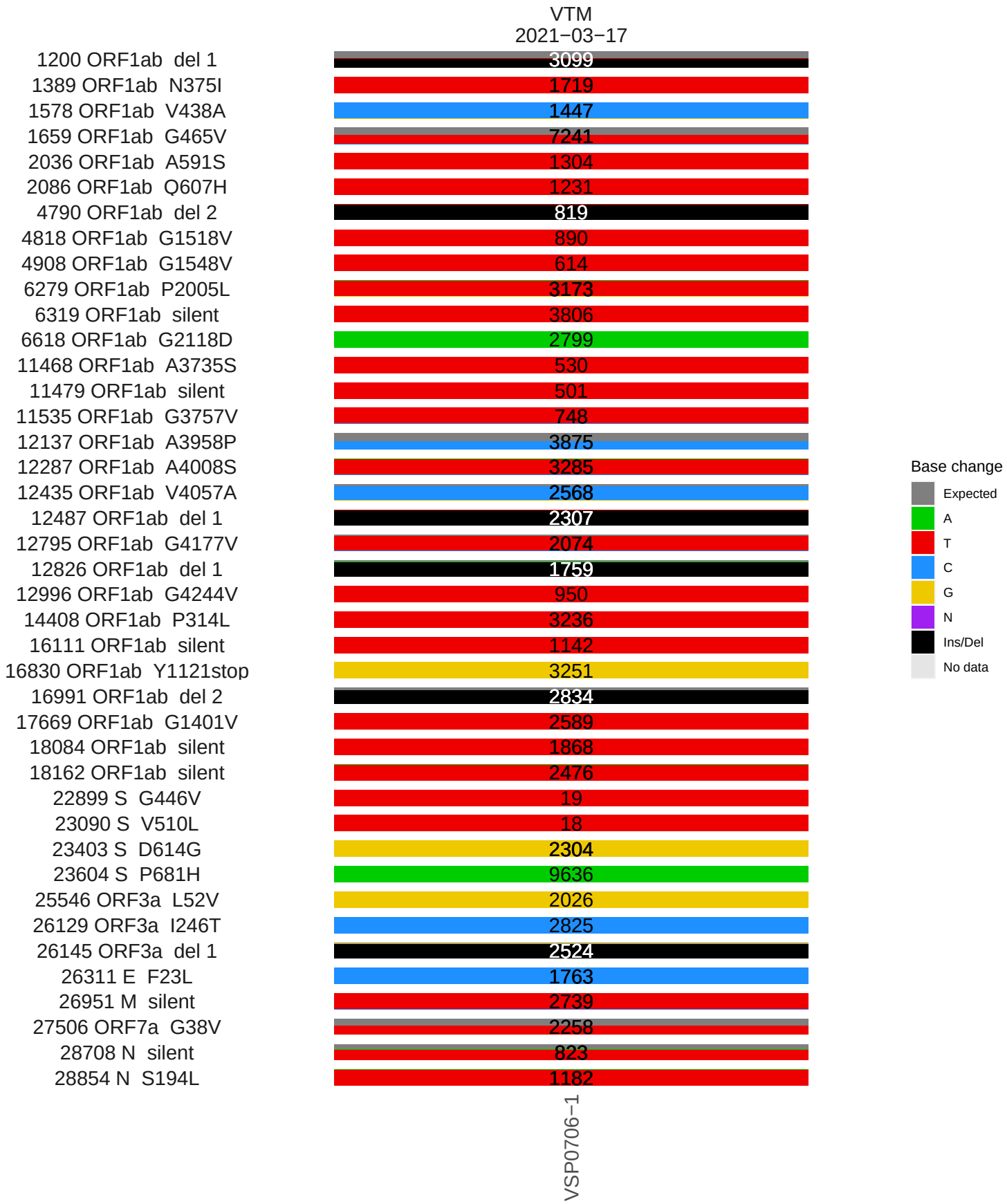
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0706-1	single experiment	NA	VTM	2021-03-17	1.81	NA	55.2%	54.4%

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.

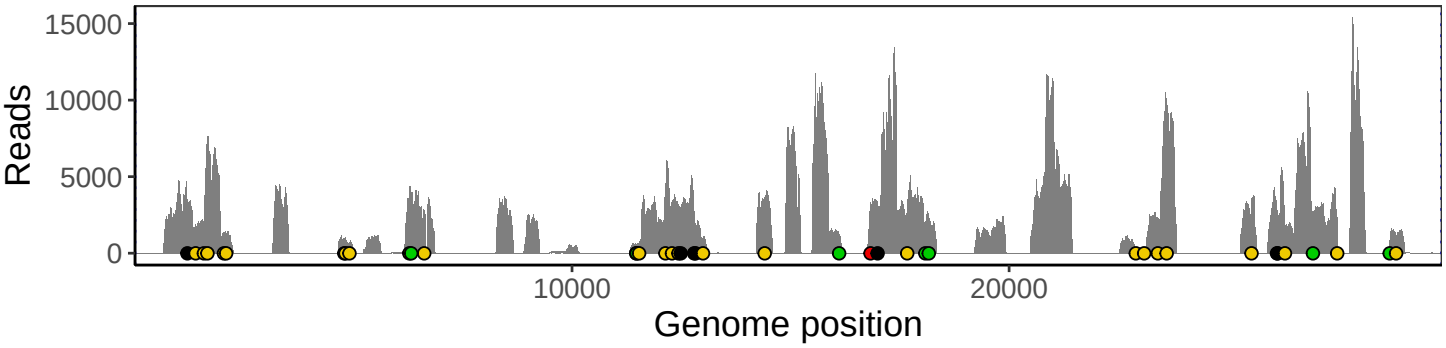



VTM
2021−03−17
1200 ORF1ab del 1
1389 ORF1ab N375I
1578 ORF1ab V438A
1659 ORF1ab G465V
2036 ORF1ab A591S
2086 ORF1ab Q607H
4790 ORF1ab del 2
4818 ORF1ab G1518V
4908 ORF1ab G1548V
6279 ORF1ab P2005L
6319 ORF1ab silent
6618 ORF1ab G2118D
11468 ORF1ab A3735S
11479 ORF1ab silent
11535 ORF1ab G3757V
12137 ORF1ab A3958P
12287 ORF1ab A4008S
12435 ORF1ab V4057A
12487 ORF1ab del 1
12795 ORF1ab G4177V
12826 ORF1ab del 1
12996 ORF1ab G4244V
14408 ORF1ab P314L
16111 ORF1ab silent
16830 ORF1ab Y1121stop
16991 ORF1ab del 2
17669 ORF1ab G1401V
18084 ORF1ab silent
18162 ORF1ab silent
22899 S G446V
23090 S V510L
23403 S D614G
23604 S P681H
25546 ORF3a L52V
26129 ORF3a I246T
26145 ORF3a del 1
26311 E F23L
26951 M silent
27506 ORF7a G38V
28708 N silent
28854 N S194L
3099
1719
1447
7241
1304
1231
819
890
614
3173
3806
2799
530
501
748
3875
3285
2568
2307
2074
1759
950
3236
1142
3251
2834
2589
1868
2476
19
18
2304
9636
2026
2825
2524
1763
2739
2258
823
1182
VSP0706−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

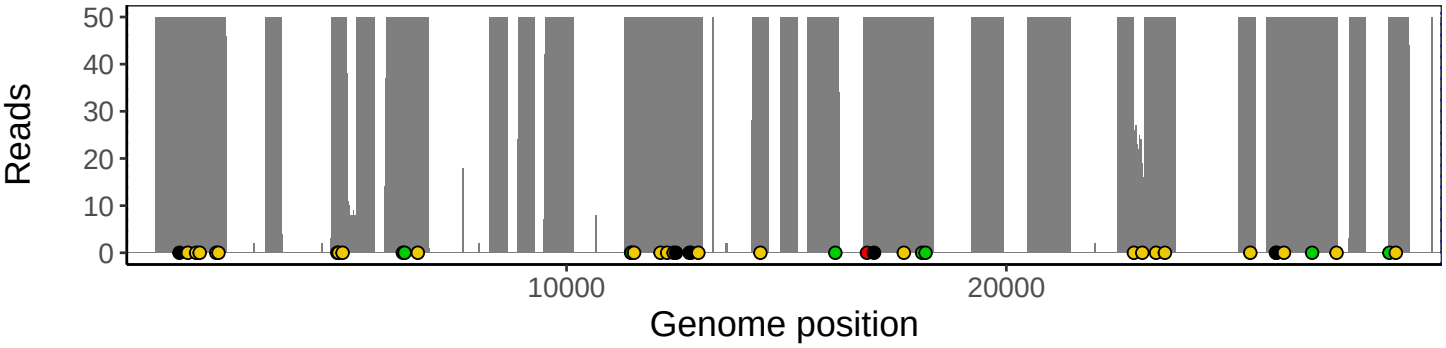
Analyses of individual experiments and composite results

VSP0706-1 | 2021-03-17 | VTM | H2103110830 | genomes | single experiment

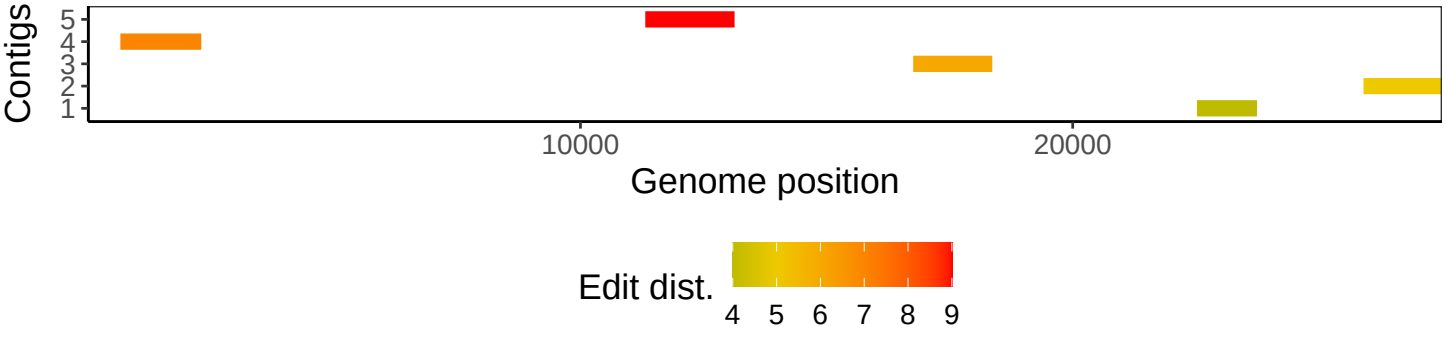
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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