COVID-19 subject H2103090778

2021-04-17

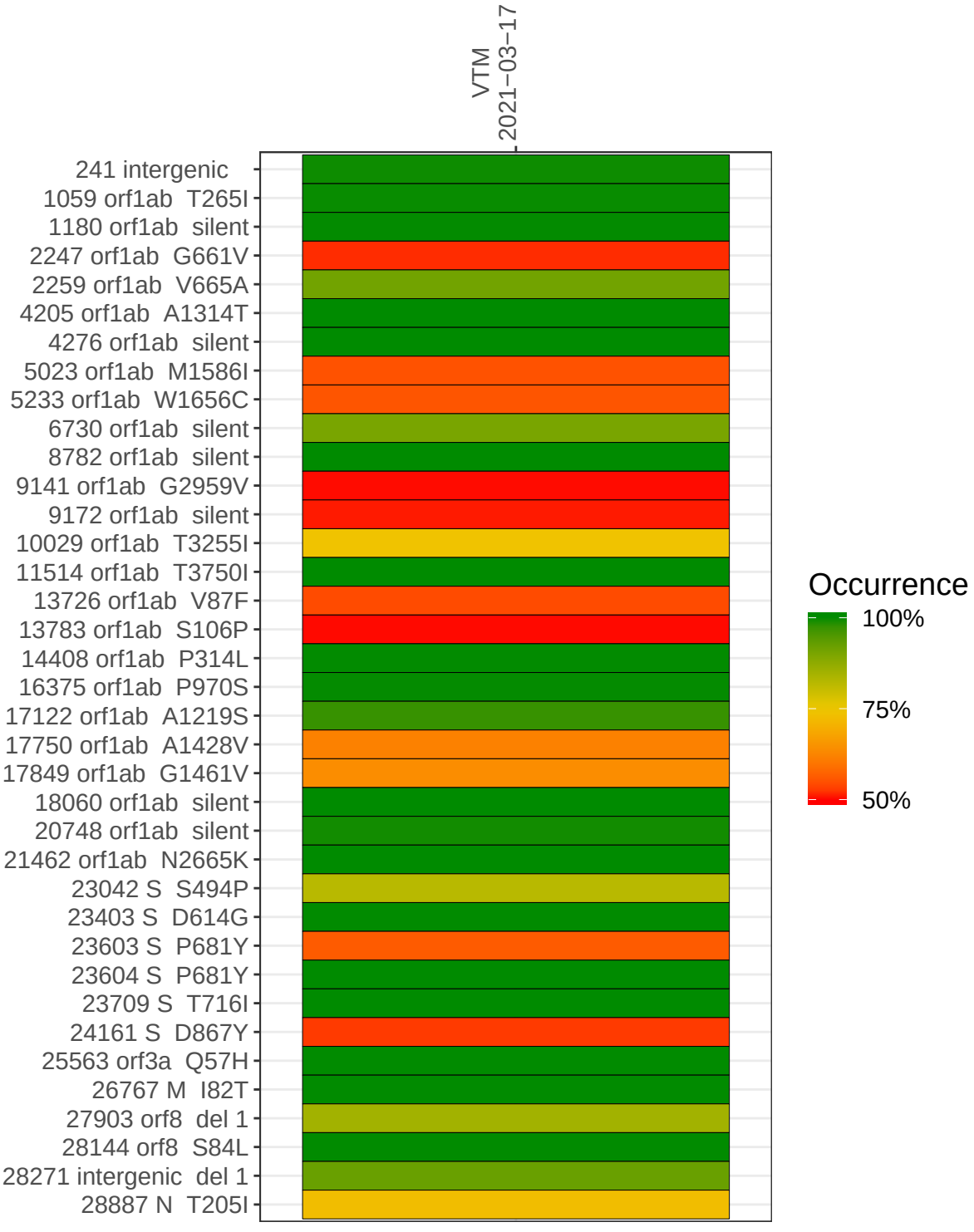
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

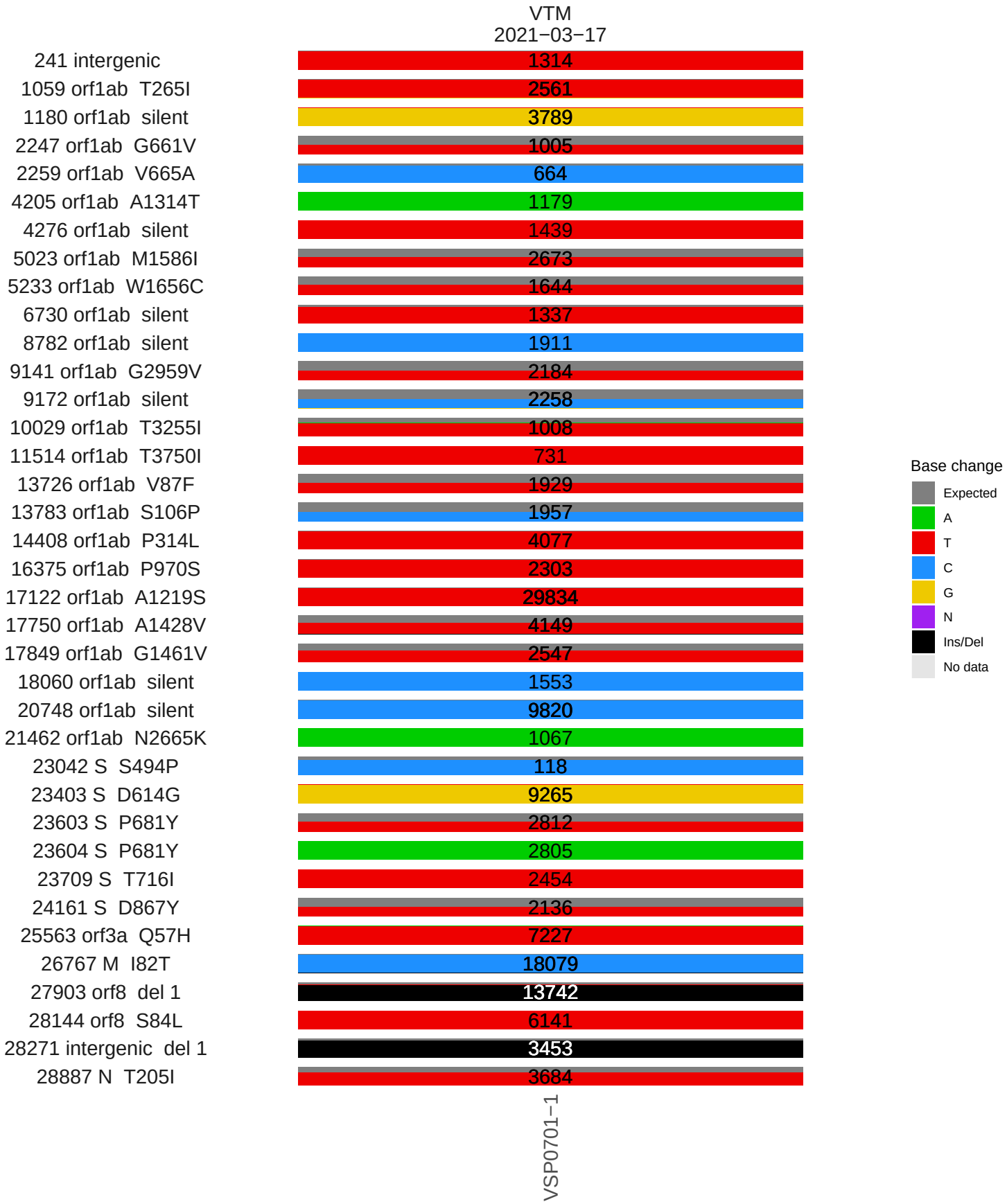
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0701-1	single experiment	NA	VTM	2021-03-17	26.57	B.1.575	97.6%	97.6%

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.

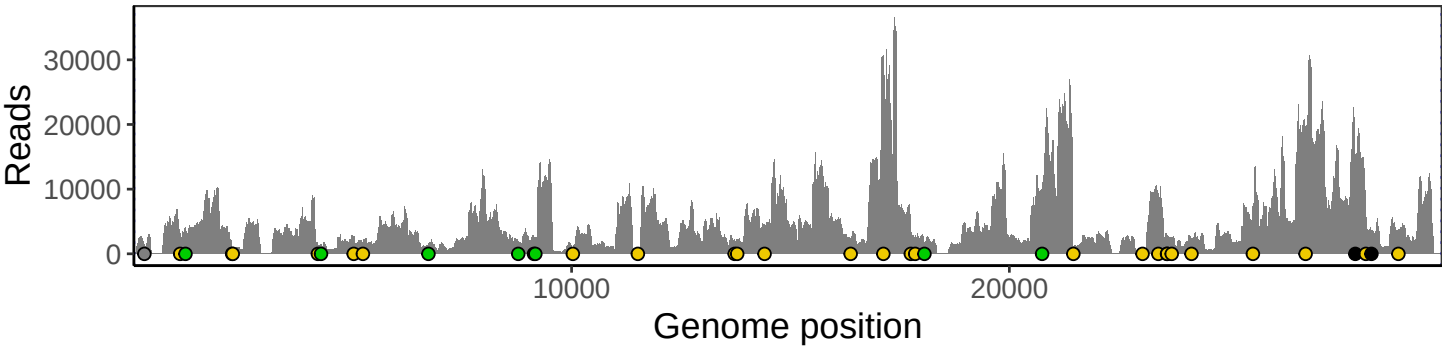



VTM
2021−03−17
241 intergenic
1059 orf1ab T265I
1180 orf1ab silent
2247 orf1ab G661V
2259 orf1ab V665A
4205 orf1ab A1314T
4276 orf1ab silent
5023 orf1ab M1586I
5233 orf1ab W1656C
6730 orf1ab silent
8782 orf1ab silent
9141 orf1ab G2959V
9172 orf1ab silent
10029 orf1ab T3255I
11514 orf1ab T3750I
13726 orf1ab V87F
13783 orf1ab S106P
14408 orf1ab P314L
16375 orf1ab P970S
17122 orf1ab A1219S
17750 orf1ab A1428V
17849 orf1ab G1461V
18060 orf1ab silent
20748 orf1ab silent
21462 orf1ab N2665K
23042 S S494P
23403 S D614G
23603 S P681Y
23604 S P681Y
23709 S T716I
24161 S D867Y
25563 orf3a Q57H
26767 M I82T
27903 orf8 del 1
28144 orf8 S84L
28271 intergenic del 1
28887 N T205I
1314
2561
3789
1005
664
1179
1439
2673
1644
1337
1911
2184
2258
1008
731
1929
1957
4077
2303
29834
4149
2547
1553
9820
1067
118
9265
2812
2805
2454
2136
7227
18079
13742
6141
3453
3684
VSP0701−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

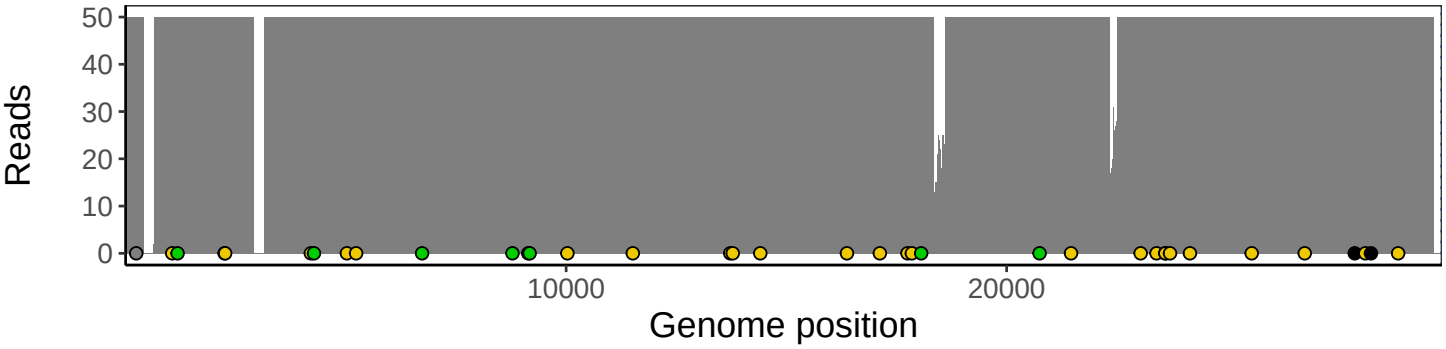
Analyses of individual experiments and composite results

VSP0701-1 | 2021-03-17 | VTM | H2103090778 | genomes | single experiment

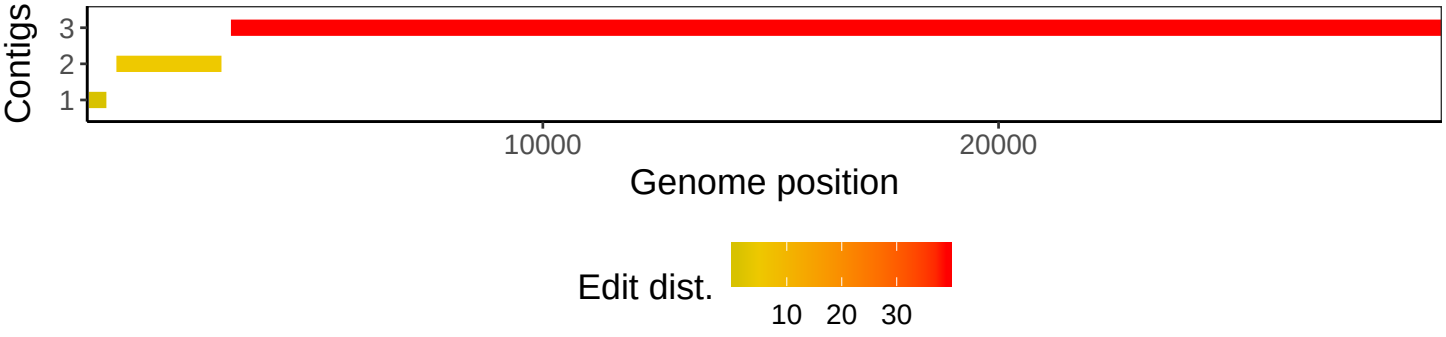
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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