COVID-19 subject S-210128-01783

2021-04-17

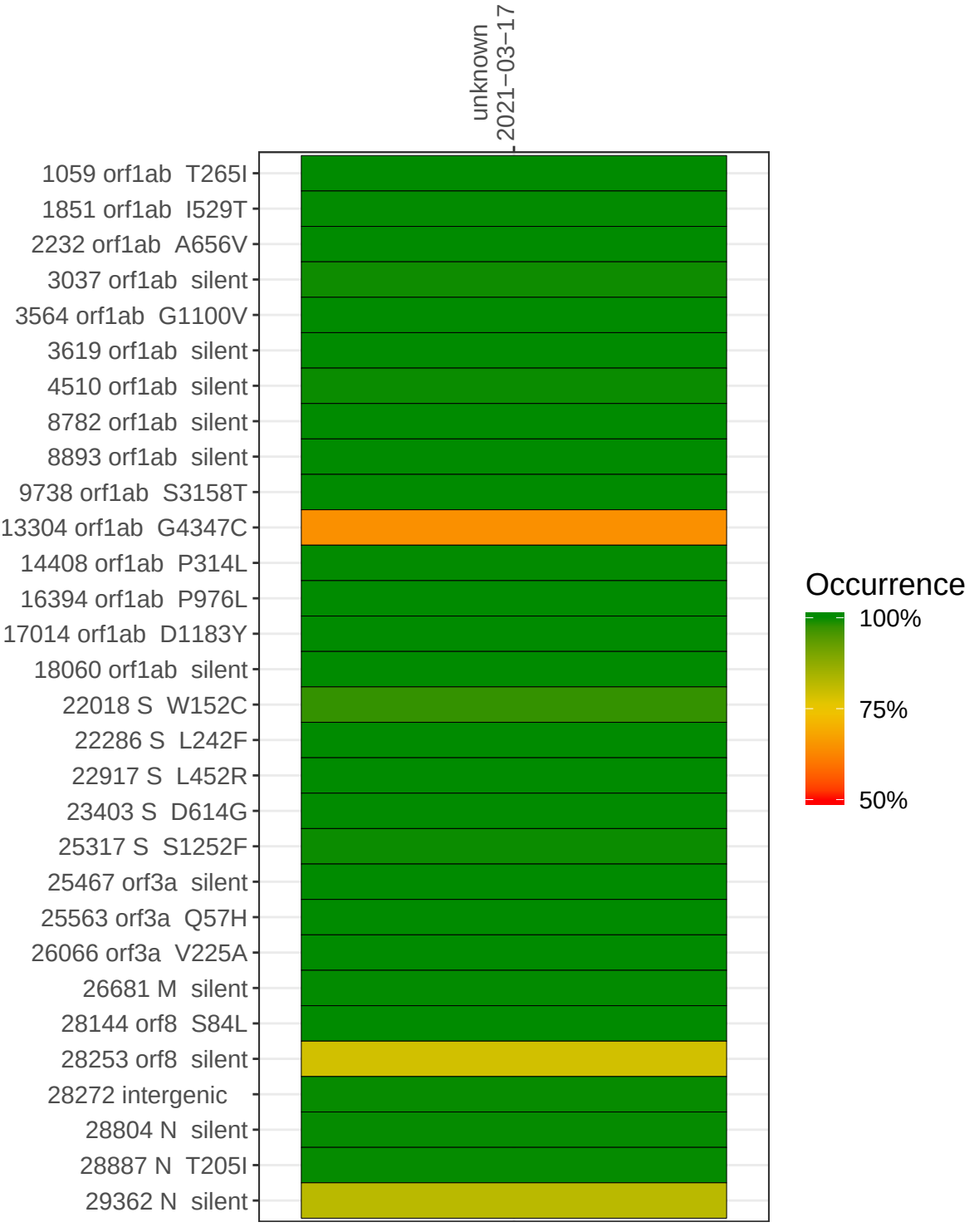
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

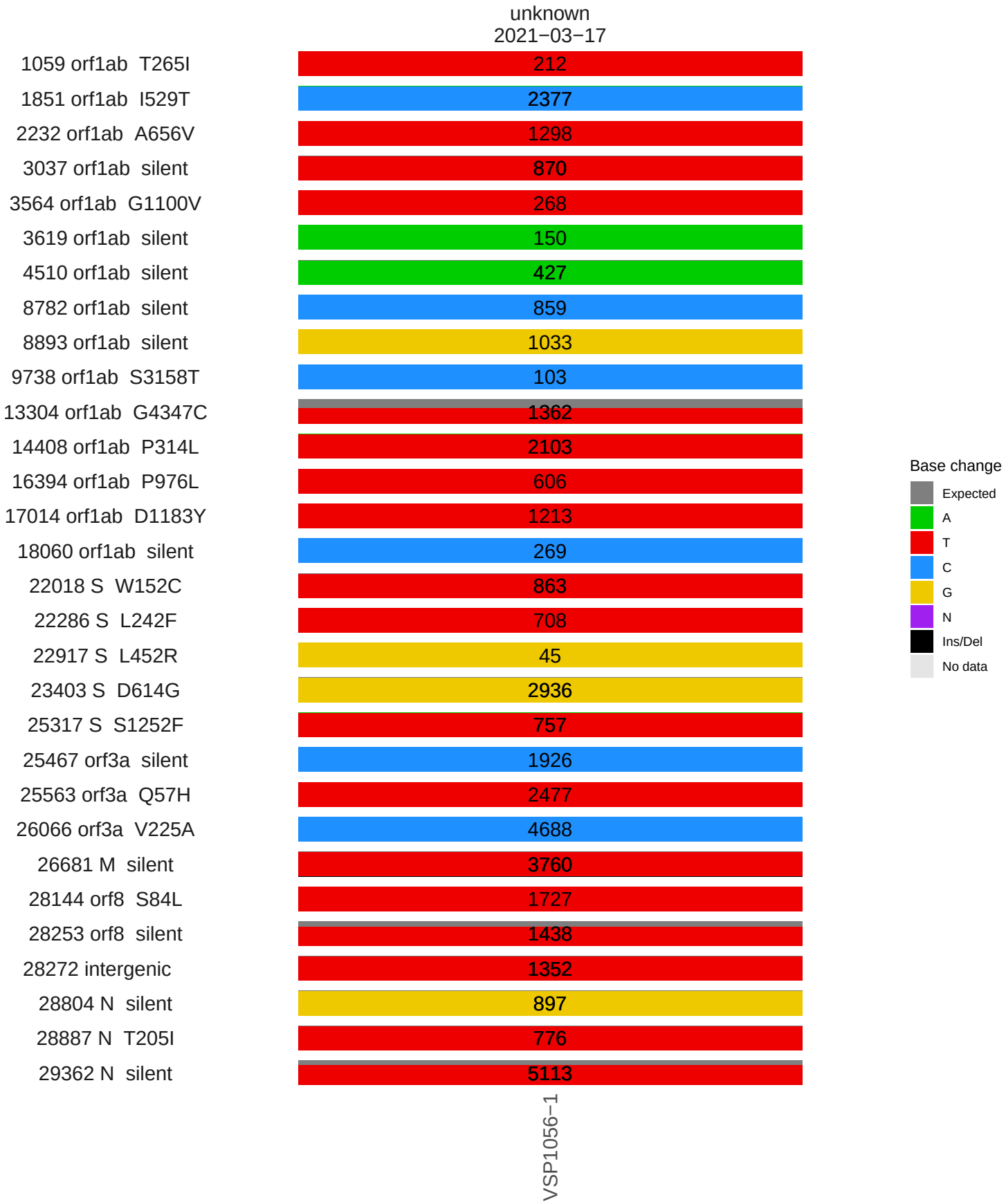
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1056-1	single experiment	NA	unknown	2021-03-17	6.48	NA	92.1%	91.6%

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.

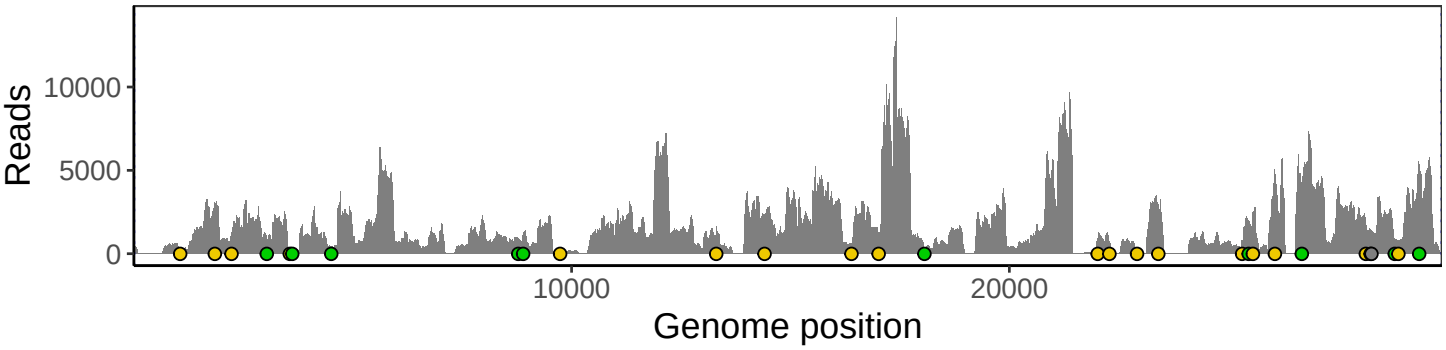



unknown
2021−03−17
1059 orf1ab T265I
1851 orf1ab I529T
2232 orf1ab A656V
3037 orf1ab silent
3564 orf1ab G1100V
3619 orf1ab silent
4510 orf1ab silent
8782 orf1ab silent
8893 orf1ab silent
9738 orf1ab S3158T
13304 orf1ab G4347C
14408 orf1ab P314L
16394 orf1ab P976L
17014 orf1ab D1183Y
18060 orf1ab silent
22018 S W152C
22286 S L242F
22917 S L452R
23403 S D614G
25317 S S1252F
25467 orf3a silent
25563 orf3a Q57H
26066 orf3a V225A
26681 M silent
28144 orf8 S84L
28253 orf8 silent
28272 intergenic
28804 N silent
28887 N T205I
29362 N silent
212
2377
1298
870
268
150
427
859
1033
103
1362
2103
606
1213
269
863
708
45
2936
757
1926
2477
4688
3760
1727
1438
1352
897
776
5113
VSP1056−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

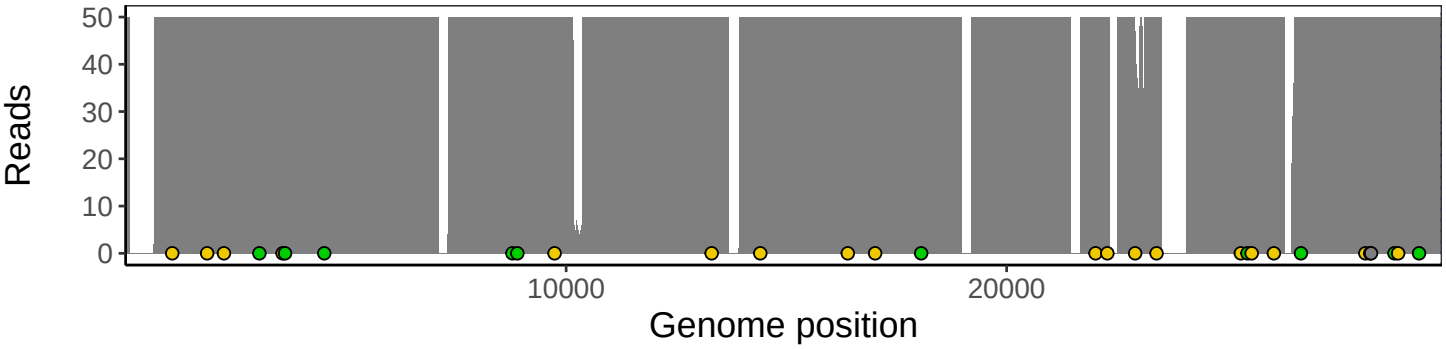
Analyses of individual experiments and composite results

VSP1056-1 | 2021-03-17 | unknown | S-210128-01783 | genomes | single experiment

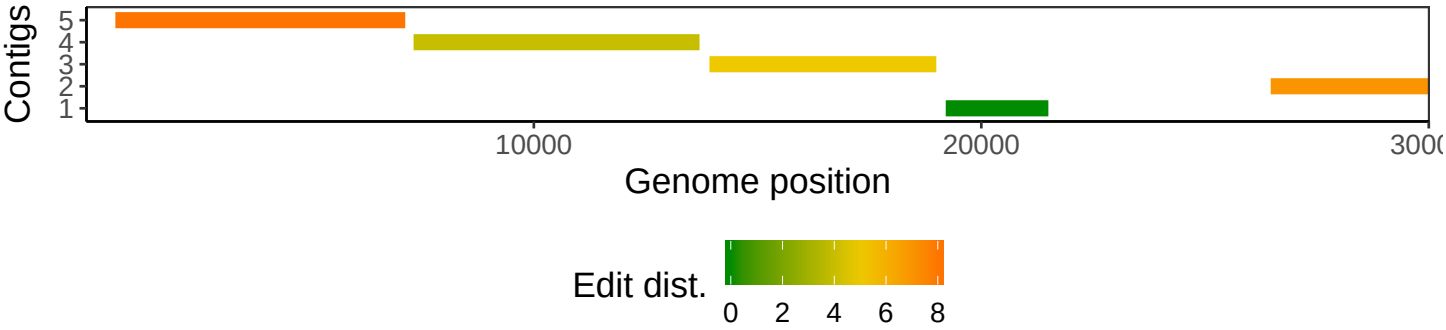
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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