COVID-19 subject S−210122−02276

2021-05-05

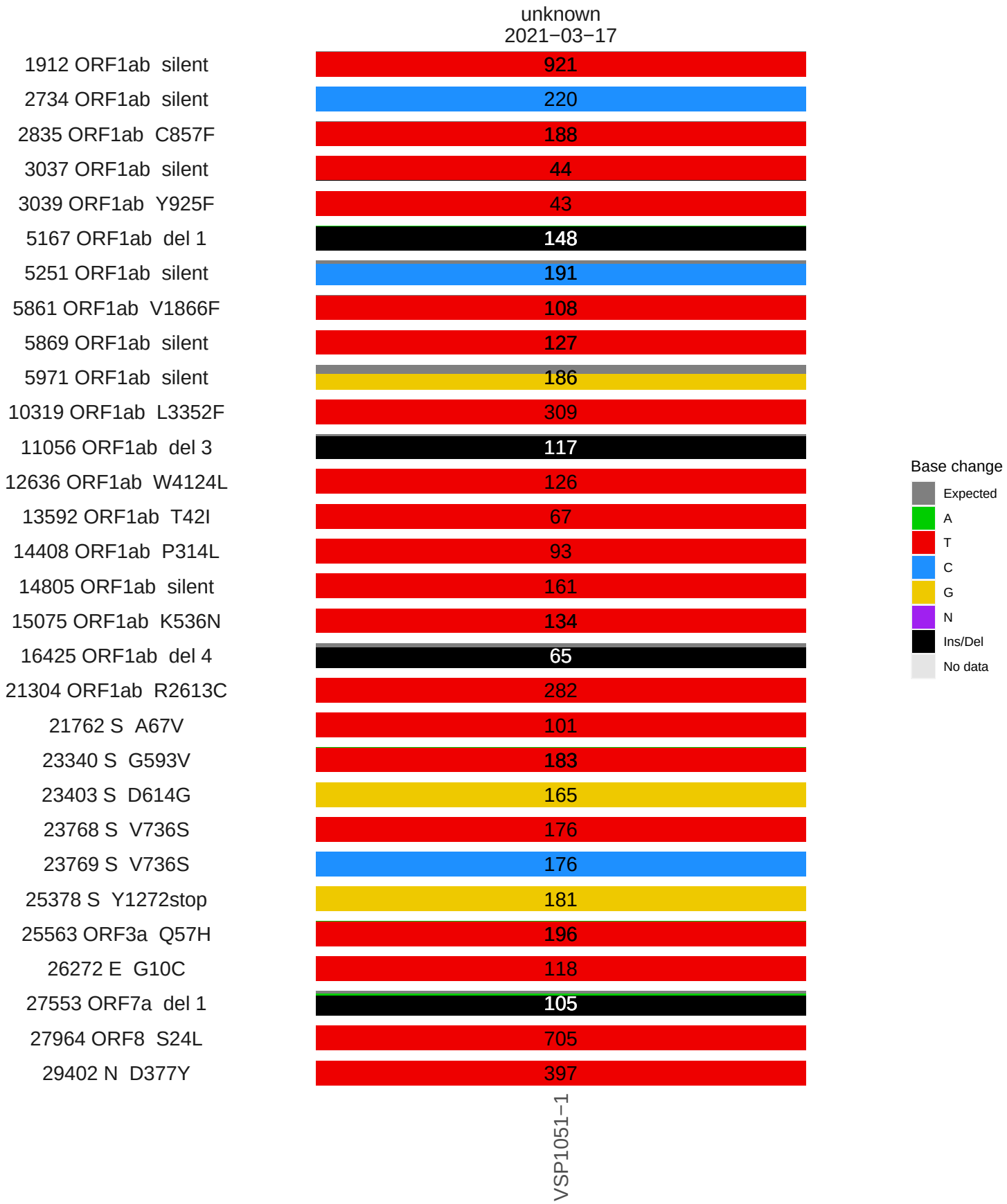
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1051-1	single experiment	NA	unknown	2021-03-17	1.96	NA	57.2%	55.9%

Variants shared across samples

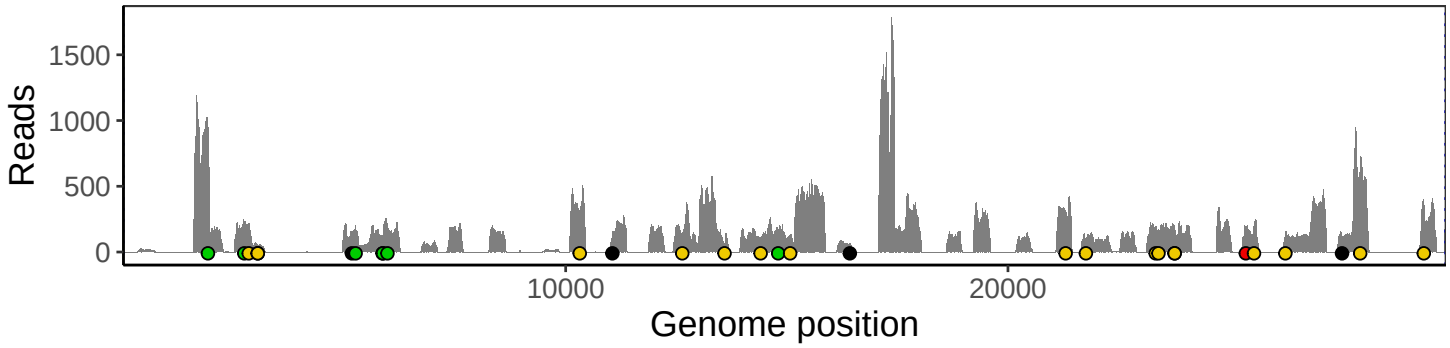
The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/Wuhan-Hu-1) are shared across subject samples where the percent variance is colored. Variants are called if a variant position is covered by 5 or more reads, the alternative base is found in > 50% of 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.


unknown
2021−03−17
1912 ORF1ab silent
2734 ORF1ab silent
2835 ORF1ab C857F
3037 ORF1ab silent
3039 ORF1ab Y925F
5167 ORF1ab del 1
5251 ORF1ab silent
5861 ORF1ab V1866F
5869 ORF1ab silent
5971 ORF1ab silent
10319 ORF1ab L3352F
11056 ORF1ab del 3
12636 ORF1ab W4124L
13592 ORF1ab T42I
14408 ORF1ab P314L
14805 ORF1ab silent
15075 ORF1ab K536N
16425 ORF1ab del 4
21304 ORF1ab R2613C
21762 S A67V
23340 S G593V
23403 S D614G
23768 S V736S
23769 S V736S
25378 S Y1272stop
25563 ORF3a Q57H
26272 E G10C
27553 ORF7a del 1
27964 ORF8 S24L
29402 N D377Y
921
220
188
44
43
148
191
108
127
186
309
117
126
67
93
161
134
65
282
101
183
165
176
176
181
196
118
105
705
397
VSP1051−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

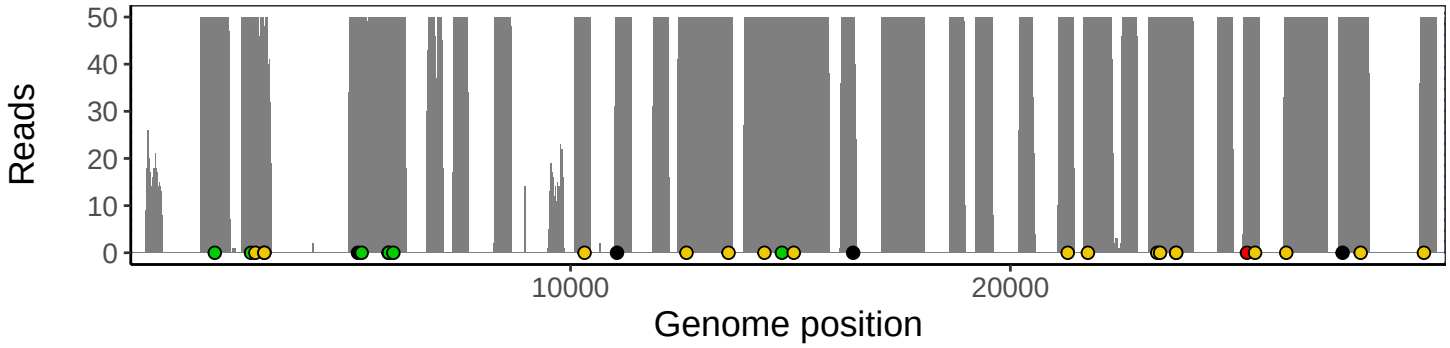
Analyses of individual experiments and composite results

VSP1051-1 | 2021-03-17 | unknown | S—210122—02276 | genomes | single experiment

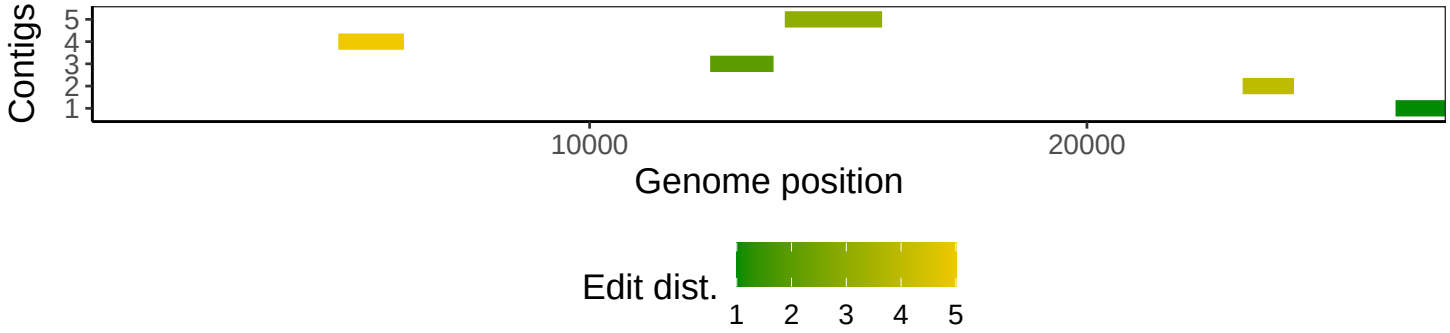
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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