COVID-19 subject H2103160769

2021-04-17

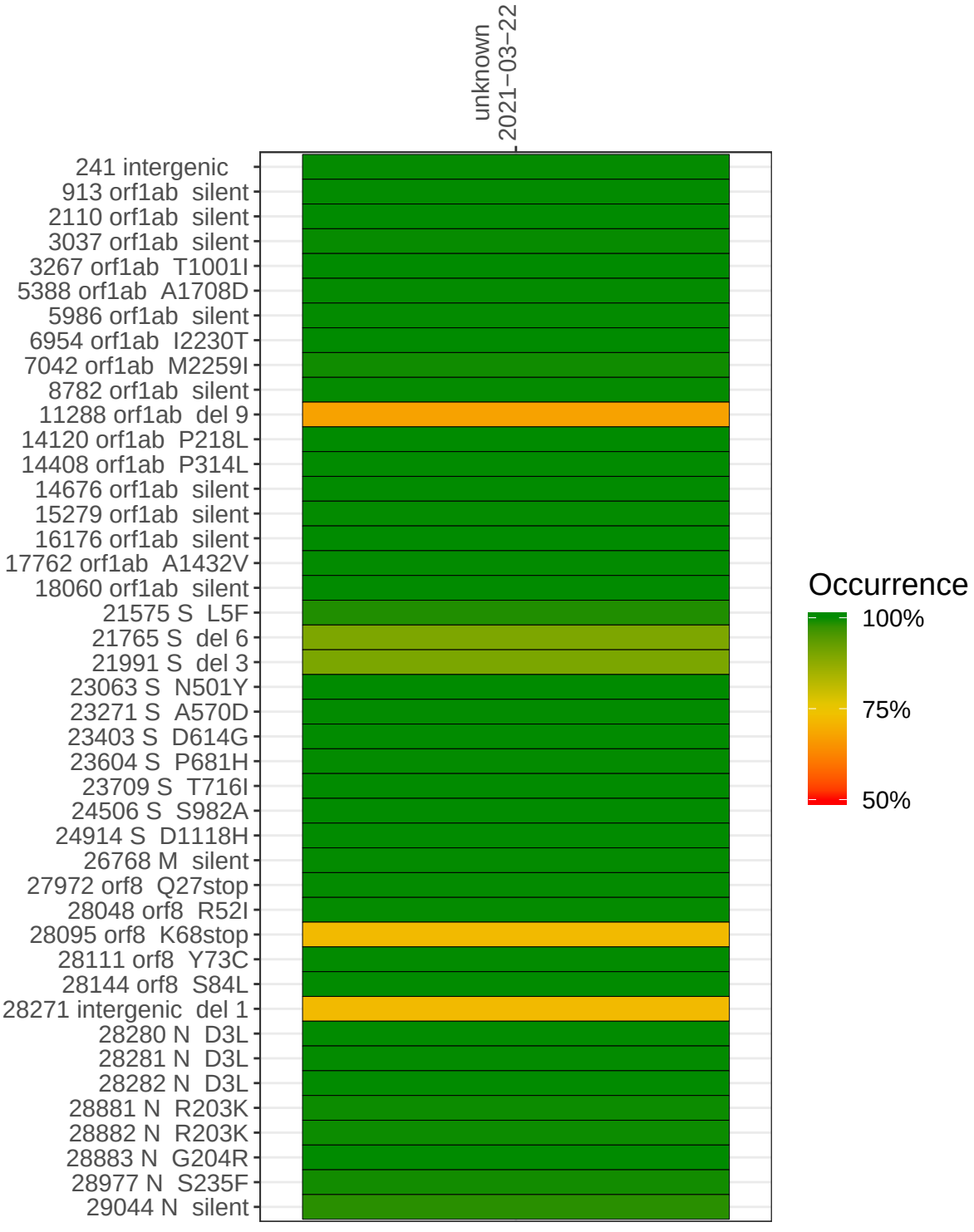
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

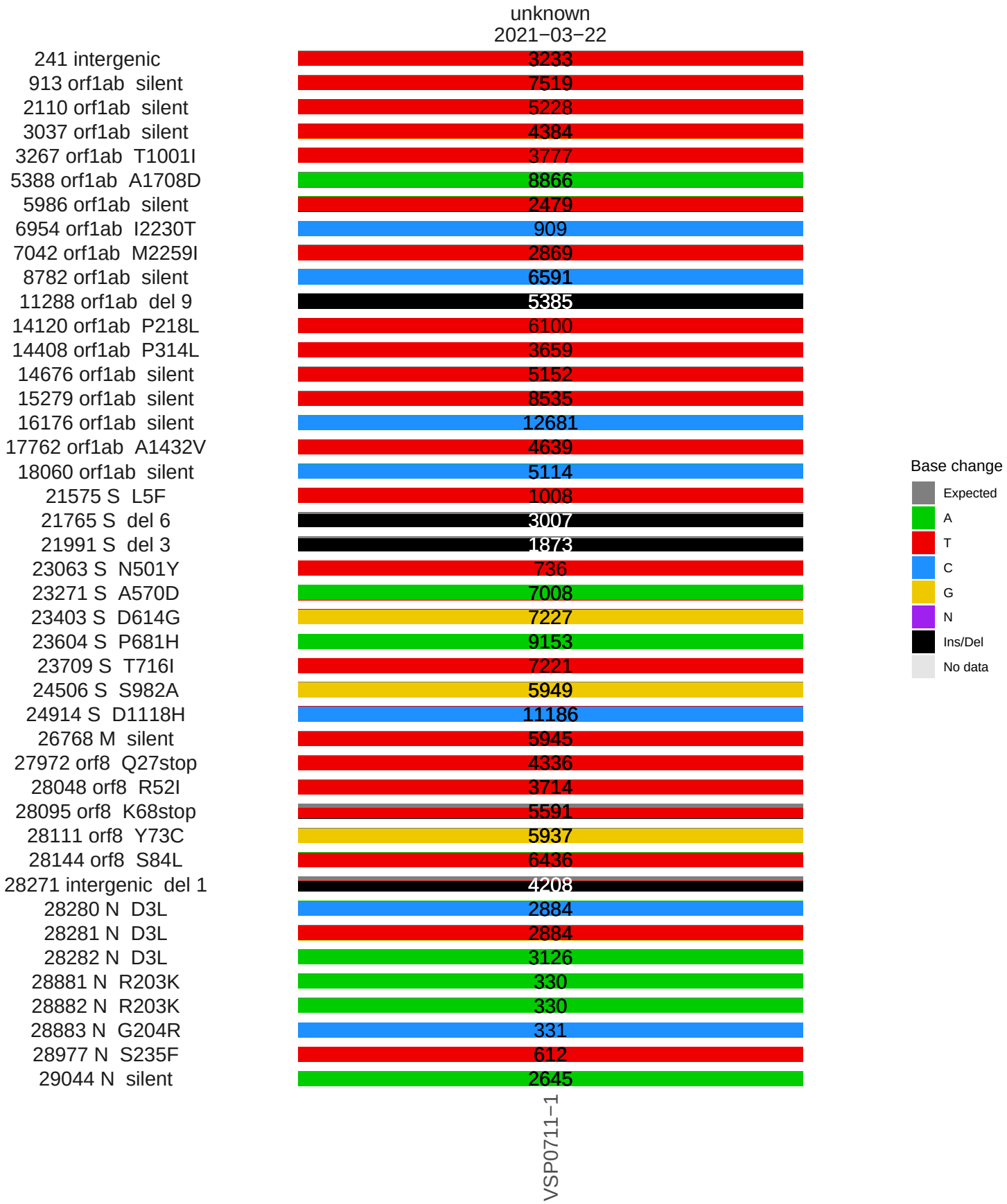
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0711-1	single experiment	NA	unknown	2021-03-22	29.80	B.1.1.7	99.9%	99.9%

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.

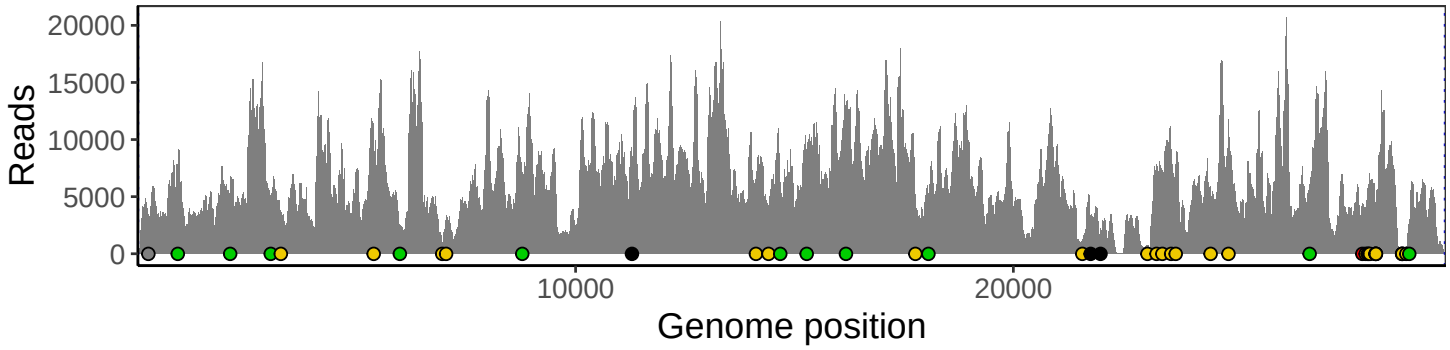



unknown
2021−03−22
241 intergenic 3233
913 orf1ab silent 7519
2110 orf1ab silent 5228
3037 orf1ab silent 4384
3267 orf1ab T1001I 3777
5388 orf1ab A1708D 8866
5986 orf1ab silent 2479
6954 orf1ab I2230T 909
7042 orf1ab M2259I 2869
8782 orf1ab silent 6591
11288 orf1ab del 9 5385
14120 orf1ab P218L 6100
14408 orf1ab P314L 3659
14676 orf1ab silent 5152
15279 orf1ab silent 8535
16176 orf1ab silent 12681
17762 orf1ab A1432V 4639
18060 orf1ab silent 5114
21575 S L5F 1008
21765 S del 6 3007
21991 S del 3 1873
23063 S N501Y 736
23271 S A570D 7008
23403 S D614G 7227
23604 S P681H 9153
23709 S T716I 7221
24506 S S982A 5949
24914 S D1118H 11186
26768 M silent 5945
27972 orf8 Q27stop 4336
28048 orf8 R52I 3714
28095 orf8 K68stop 5591
28111 orf8 Y73C 5937
28144 orf8 S84L 6436
28271 intergenic del 1 4208
28280 N D3L 2884
28281 N D3L 2884
28282 N D3L 3126
28881 N R203K 330
28882 N R203K 330
28883 N G204R 331
28977 N S235F 612
29044 N silent 2645
VSP0711−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

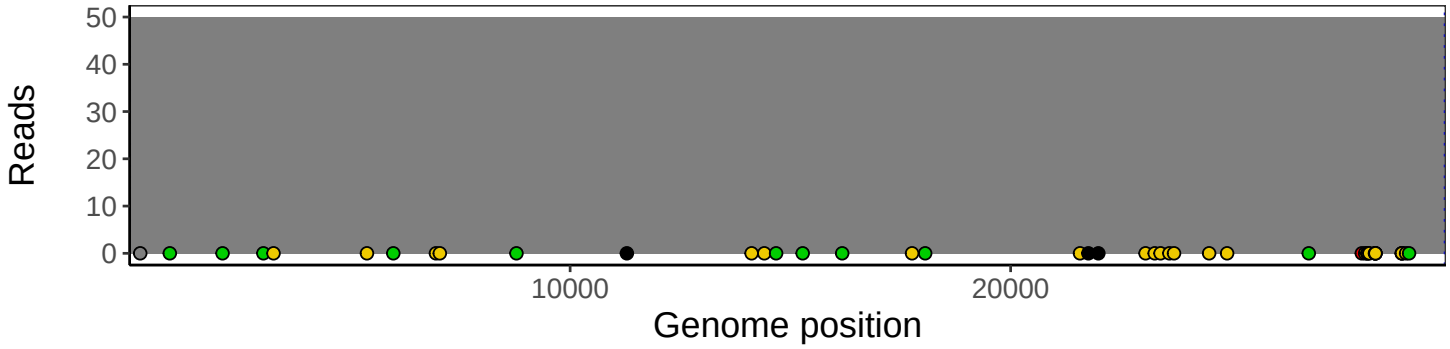
Analyses of individual experiments and composite results

VSP0711-1 | 2021-03-22 | unknown | H2103160769 | genomes | single experiment

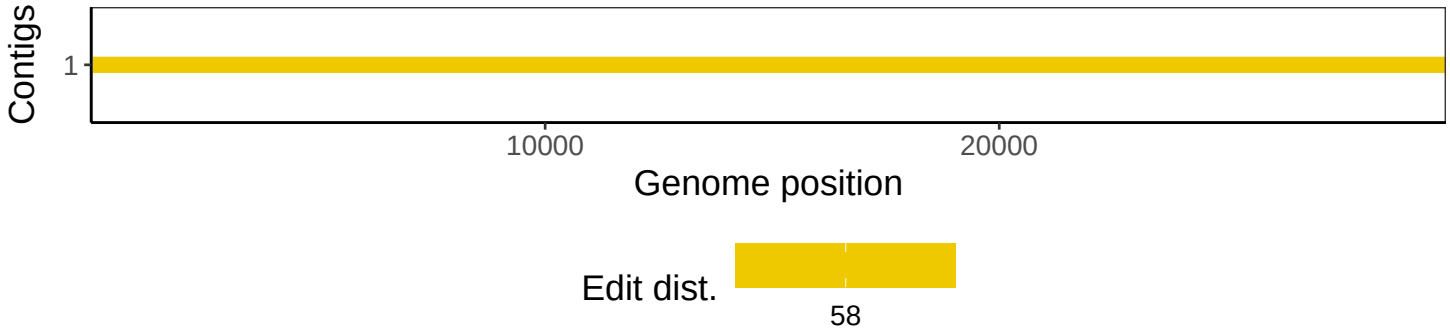
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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