COVID-19 subject UPHS-0033

2021-05-05

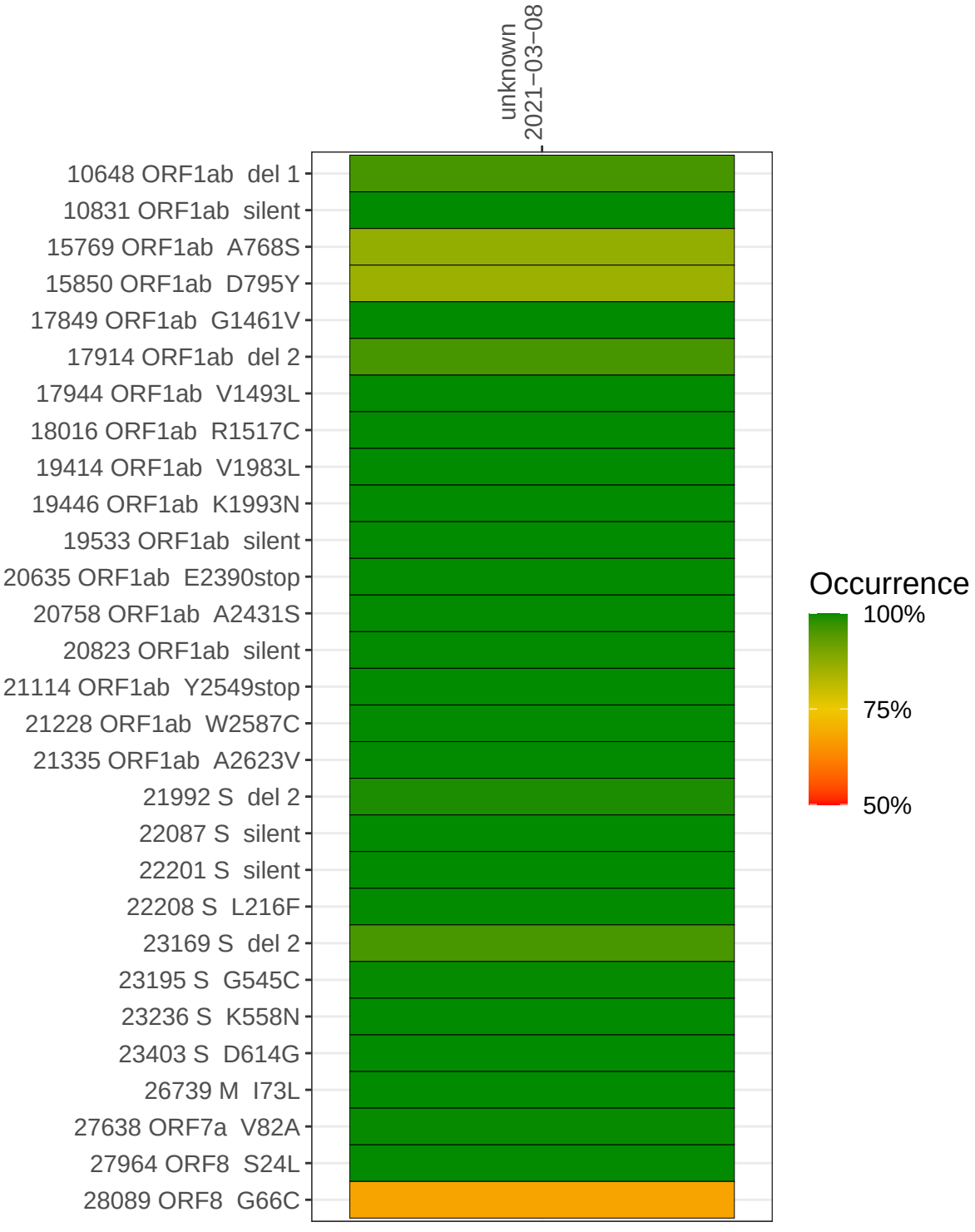
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

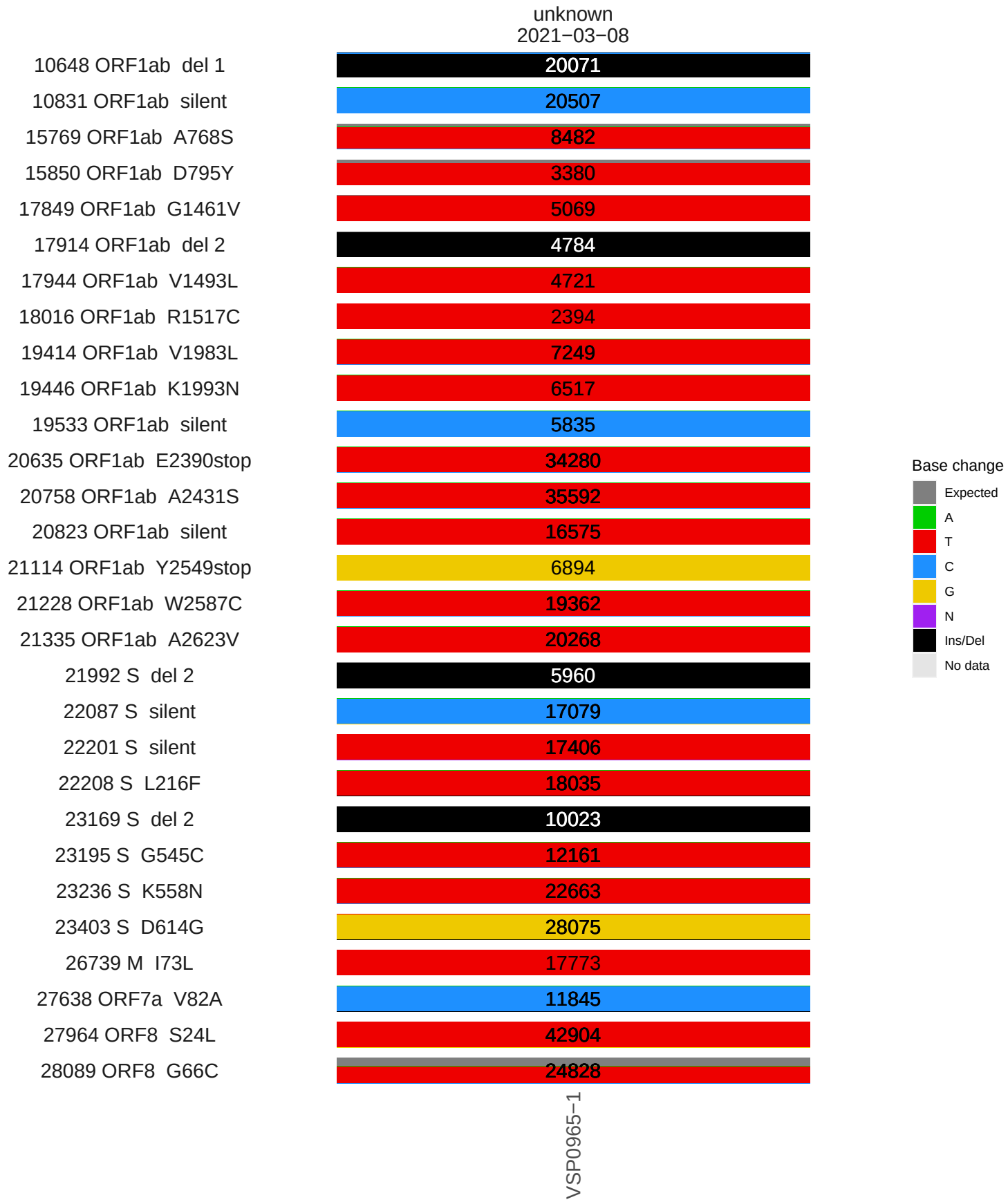
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0965-1	single experiment	NA	unknown	2021-03-08	0.71	NA	24.9%	18.7%

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.

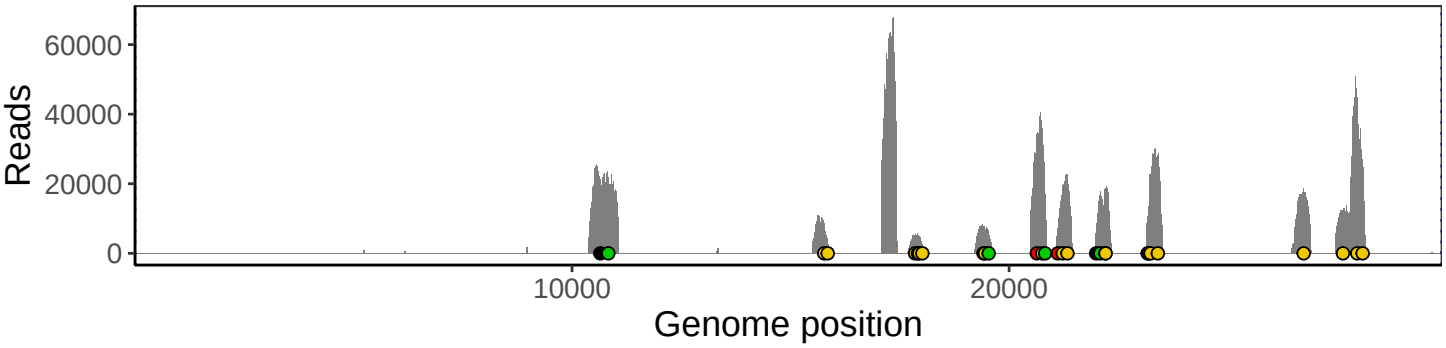



unknown
2021−03−08
10648 ORF1ab del 1
10831 ORF1ab silent
15769 ORF1ab A768S
15850 ORF1ab D795Y
17849 ORF1ab G1461V
17914 ORF1ab del 2
17944 ORF1ab V1493L
18016 ORF1ab R1517C
19414 ORF1ab V1983L
19446 ORF1ab K1993N
19533 ORF1ab silent
20635 ORF1ab E2390stop
20758 ORF1ab A2431S
20823 ORF1ab silent
21114 ORF1ab Y2549stop
21228 ORF1ab W2587C
21335 ORF1ab A2623V
21992 S del 2
22087 S silent
22201 S silent
22208 S L216F
23169 S del 2
23195 S G545C
23236 S K558N
23403 S D614G
26739 M I73L
27638 ORF7a V82A
27964 ORF8 S24L
28089 ORF8 G66C
20071
20507
8482
3380
5069
4784
4721
2394
7249
6517
5835
34280
35592
16575
6894
19362
20268
5960
17079
17406
18035
10023
12161
22663
28075
17773
11845
42904
24828
VSP0965−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

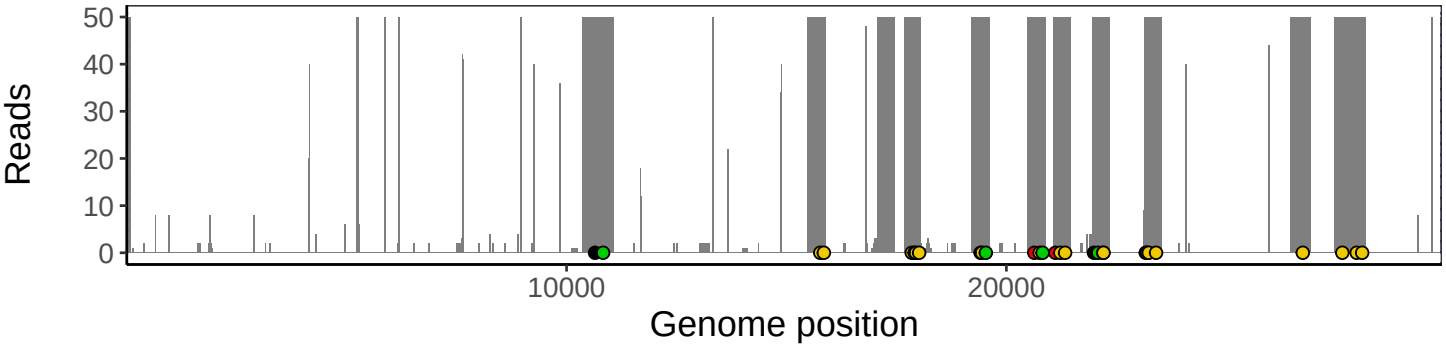
Analyses of individual experiments and composite results

VSP0965-1 | 2021-03-08 | unknown | UPHS-0033 | genomes | single experiment

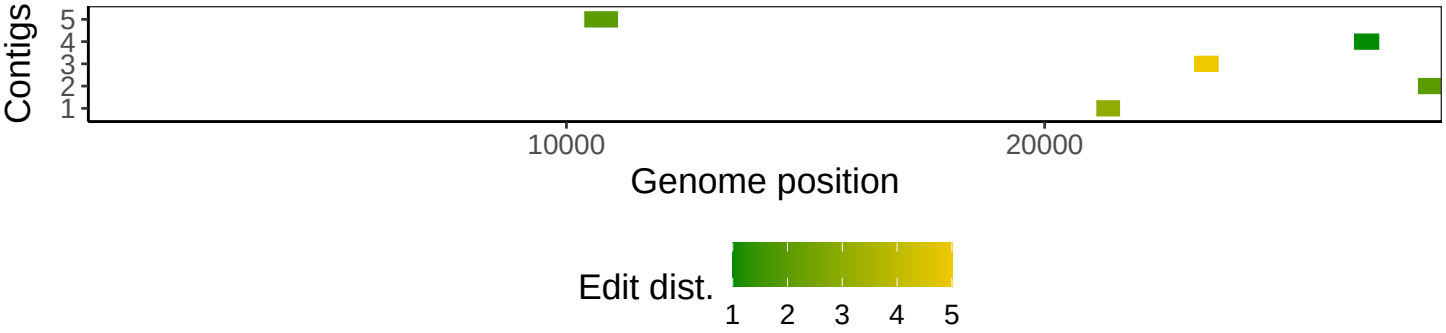
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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