

COVID-19 subject UPHS-0011

2021-06-23

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.

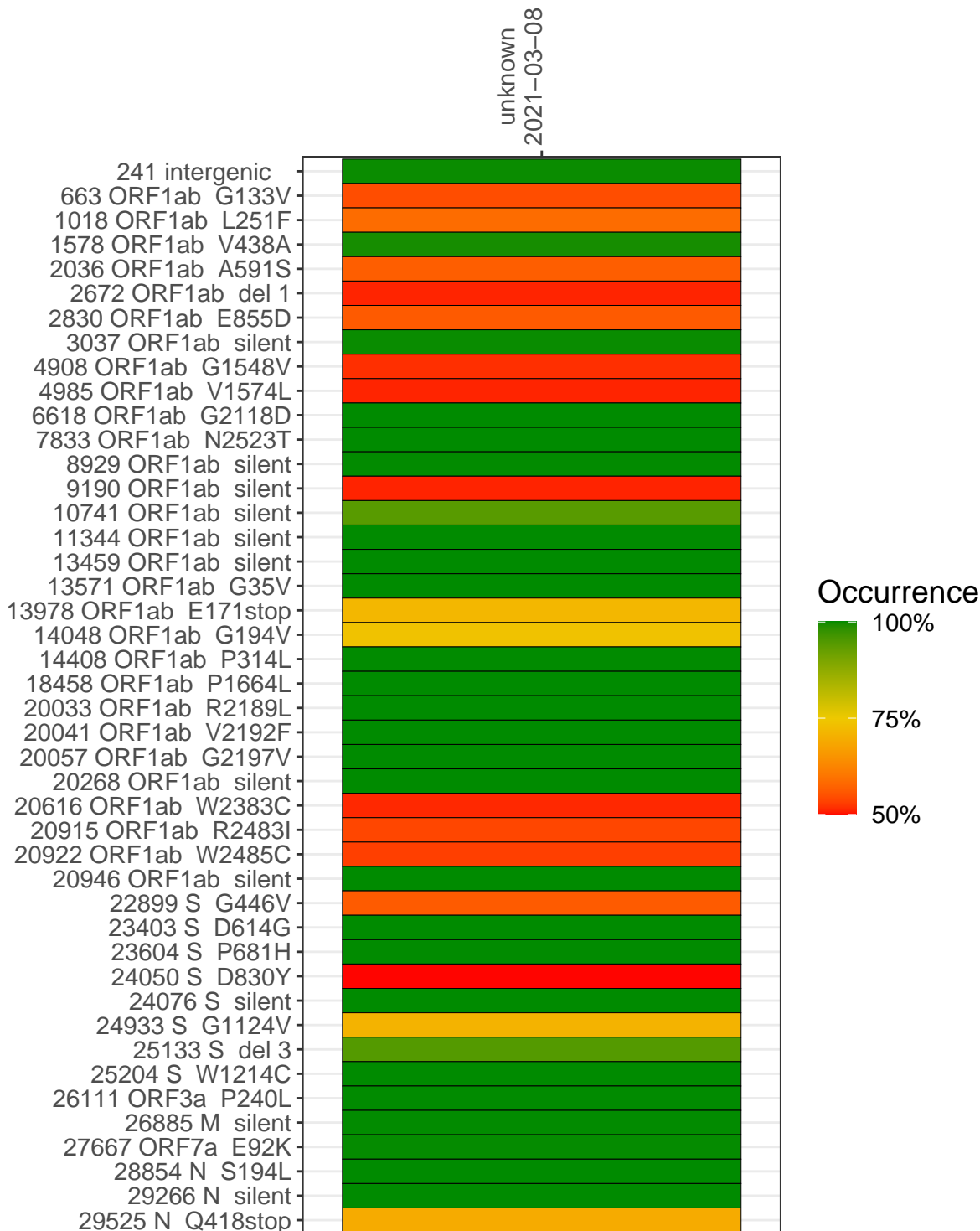
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with $> 90\%$ sequence coverage.

Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (≥ 5 reads)
VSP0944-1	single experiment	NA	unknown	2021-03-08	12.82	B.1.243	96.1%	96.0%

Variants shared across samples

The heat map below shows how variants (reference genome /home/common/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

241 intergenic	844
663 ORF1ab G133V	745
1018 ORF1ab L251F	3735
1578 ORF1ab V438A	952
2036 ORF1ab A591S	4447
2672 ORF1ab del 1	2465
2830 ORF1ab E855D	2187
3037 ORF1ab silent	958
4908 ORF1ab G1548V	1717
4985 ORF1ab V1574L	561
6618 ORF1ab G2118D	1323
7833 ORF1ab N2523T	13452
8929 ORF1ab silent	5052
9190 ORF1ab silent	1897
10741 ORF1ab silent	2297
11344 ORF1ab silent	16281
13459 ORF1ab silent	705
13571 ORF1ab G35V	391
13978 ORF1ab E171stop	724
14048 ORF1ab G194V	1096
14408 ORF1ab P314L	7303
18458 ORF1ab P1664L	1723
20033 ORF1ab R2189L	809
20041 ORF1ab V2192F	785
20057 ORF1ab G2197V	714
20268 ORF1ab silent	1146
20616 ORF1ab W2383C	1501
20915 ORF1ab R2483I	3127
20922 ORF1ab W2485C	3131
20946 ORF1ab silent	3236
22899 S G446V	32
23403 S D614G	2932
23604 S P681H	3848
24050 S D830Y	645
24076 S silent	533
24933 S G1124V	7671
25133 S del 3	728
25204 S W1214C	602
26111 ORF3a P240L	9893
26885 M silent	11512
27667 ORF7a E92K	5004
28854 N S194L	173
29266 N silent	2469
29525 N Q418stop	2331

Base change

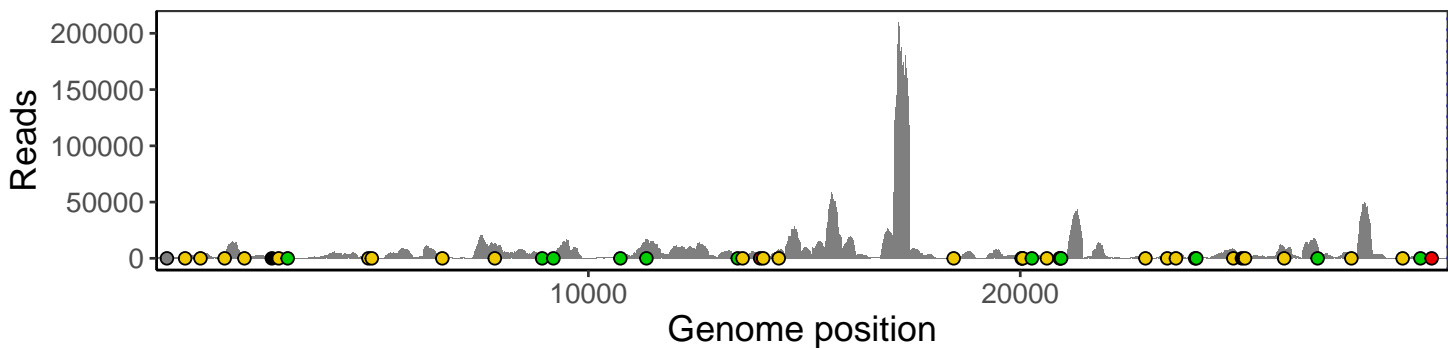


VSP0944-1

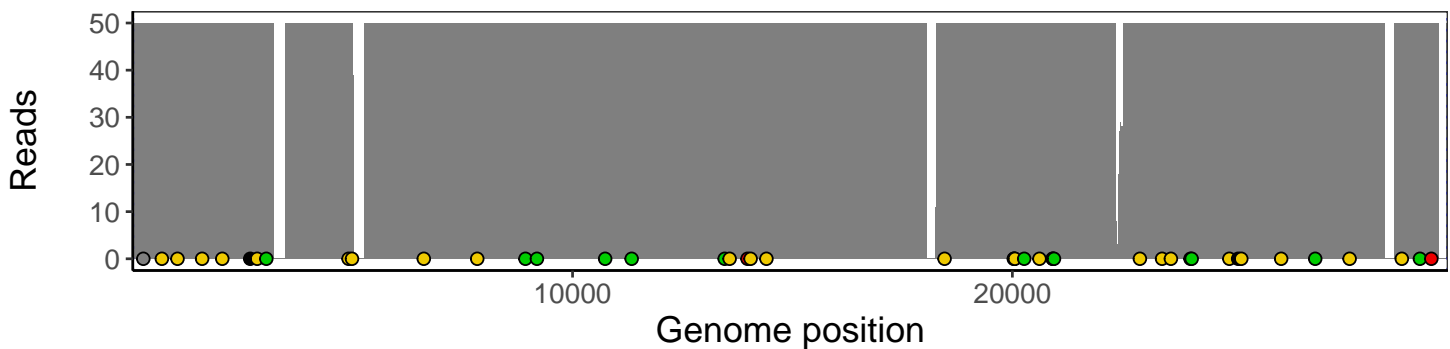
Analyses of individual experiments and composite results

VSP0944-1 | 2021-03-08 | unknown | UPHS-0011 | genomes | single experiment

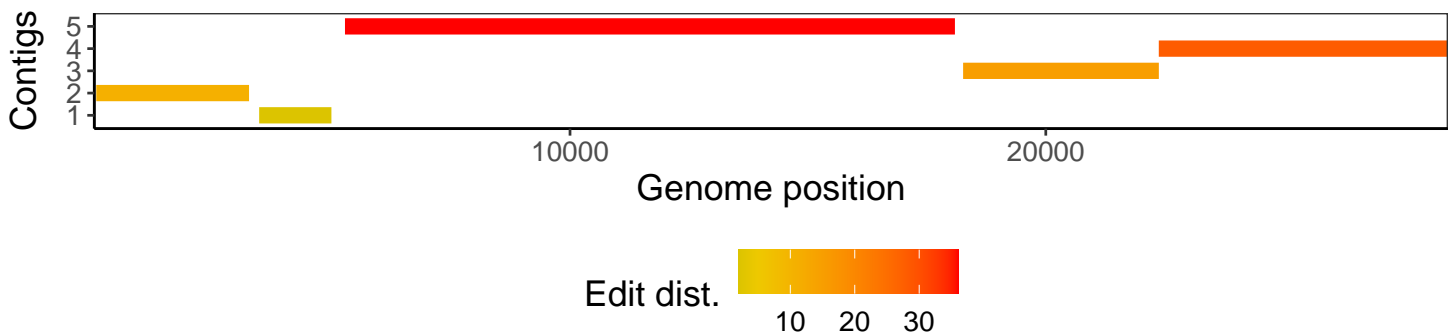
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according to variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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 htlib 1.10
bcftools	1.10.2-34-g1a12af0-dirty Using htlib 1.10.2-57-gf58a6f3
pangolin	3.1.3
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	3.3.3
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