COVID-19 subject UPHS-0011

2021-03-25

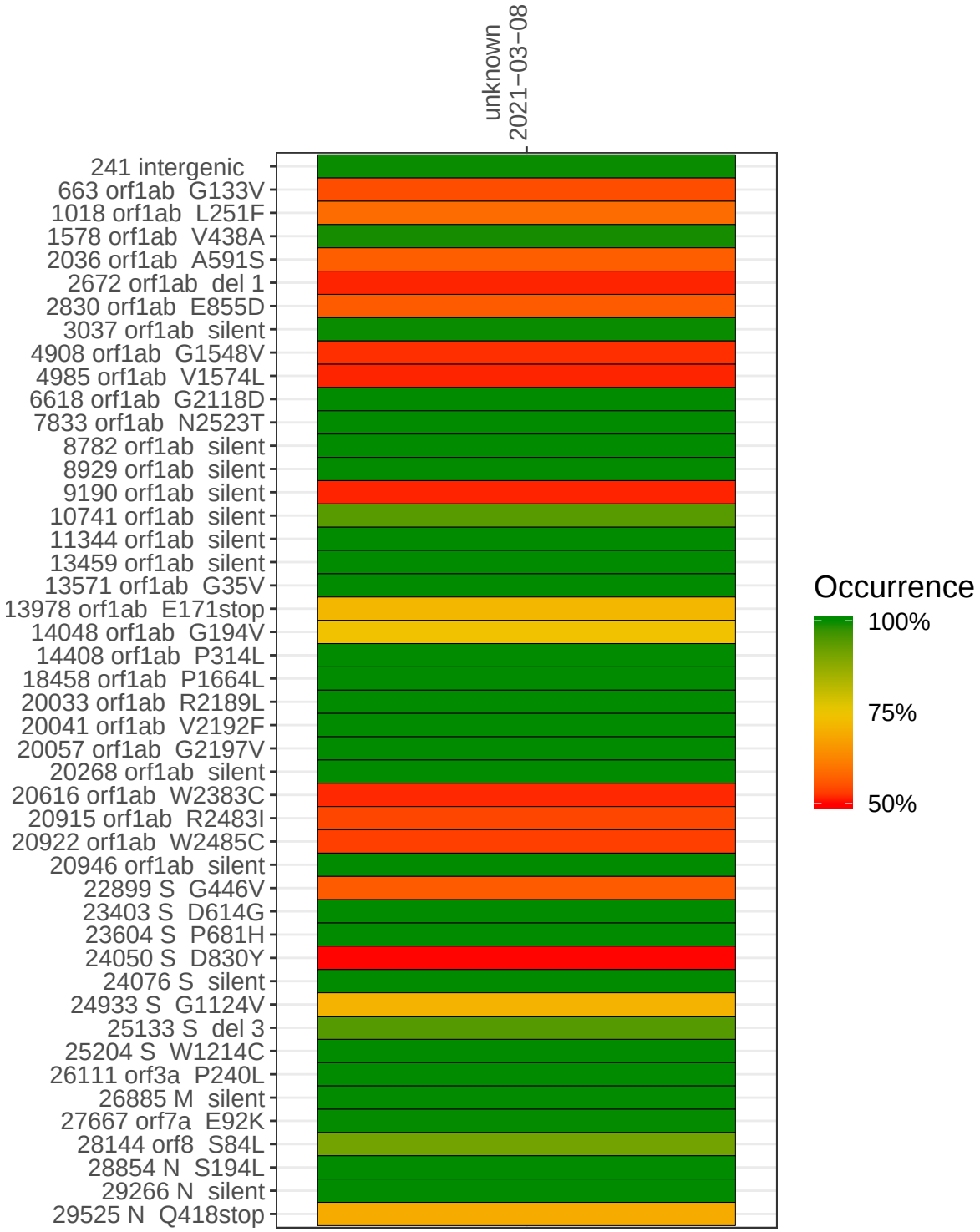
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

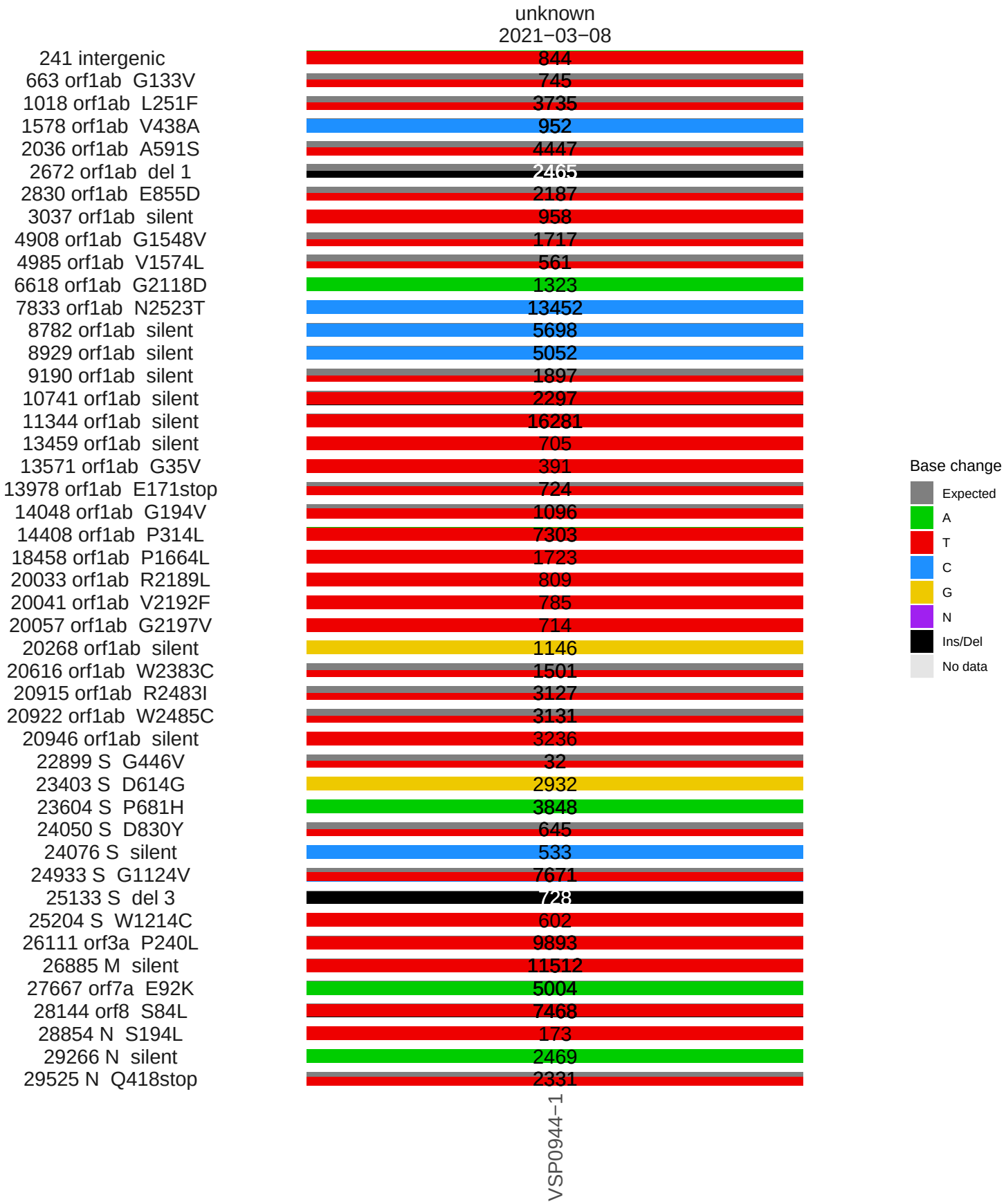
Table 1. Sample summary.

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%	95.9%

Variants shared across samples

The heat map below shows how variants (reference genome 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.

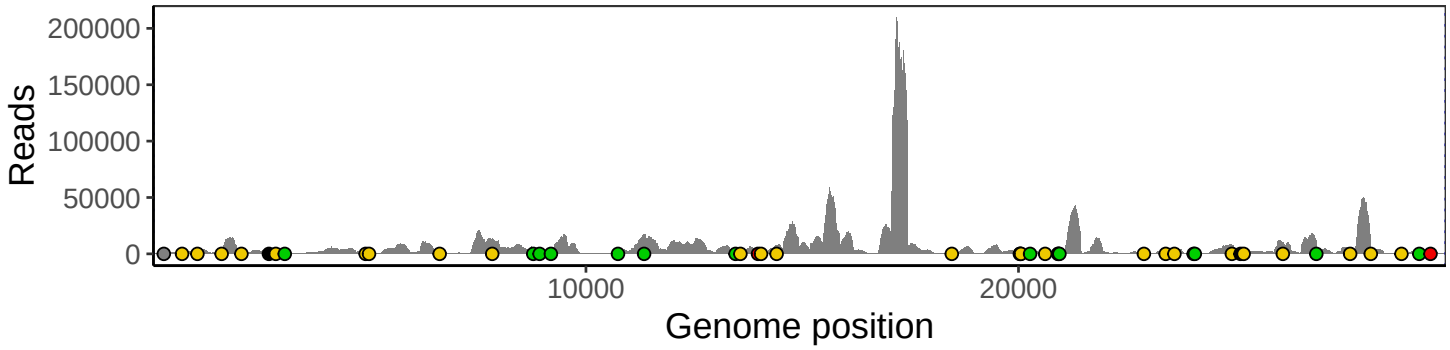



unknown
2021−03−08
241 intergenic
663 orf1ab G133V
1018 orf1ab L251F
1578 orf1ab V438A
2036 orf1ab A591S
2672 orf1ab del 1
2830 orf1ab E855D
3037 orf1ab silent
4908 orf1ab G1548V
4985 orf1ab V1574L
6618 orf1ab G2118D
7833 orf1ab N2523T
8782 orf1ab silent
8929 orf1ab silent
9190 orf1ab silent
10741 orf1ab silent
11344 orf1ab silent
13459 orf1ab silent
13571 orf1ab G35V
13978 orf1ab E171stop
14048 orf1ab G194V
14408 orf1ab P314L
18458 orf1ab P1664L
20033 orf1ab R2189L
20041 orf1ab V2192F
20057 orf1ab G2197V
20268 orf1ab silent
20616 orf1ab W2383C
20915 orf1ab R2483I
20922 orf1ab W2485C
20946 orf1ab silent
22899 S G446V
23403 S D614G
23604 S P681H
24050 S D830Y
24076 S silent
24933 S G1124V
25133 S del 3
25204 S W1214C
26111 orf3a P240L
26885 M silent
27667 orf7a E92K
28144 orf8 S84L
28854 N S194L
29266 N silent
29525 N Q418stop
844
745
3735
952
4447
2465
2187
958
1717
561
1323
13452
5698
5052
1897
2297
16281
705
391
724
1096
7303
1723
809
785
714
1146
1501
3127
3131
3236
32
2932
3848
645
533
7671
728
602
9893
11512
5004
7468
173
2469
2331
VSP0944−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

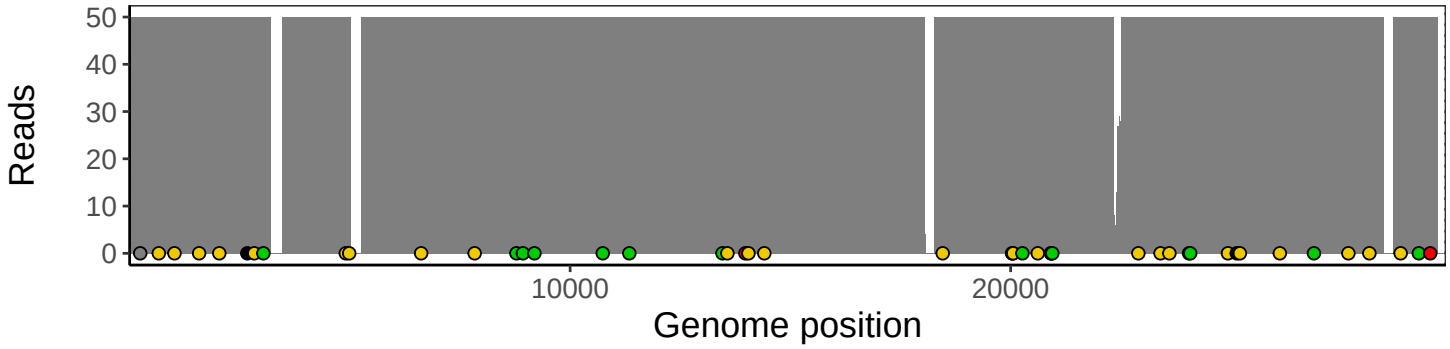
Analyses of individual experiments and composite results

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

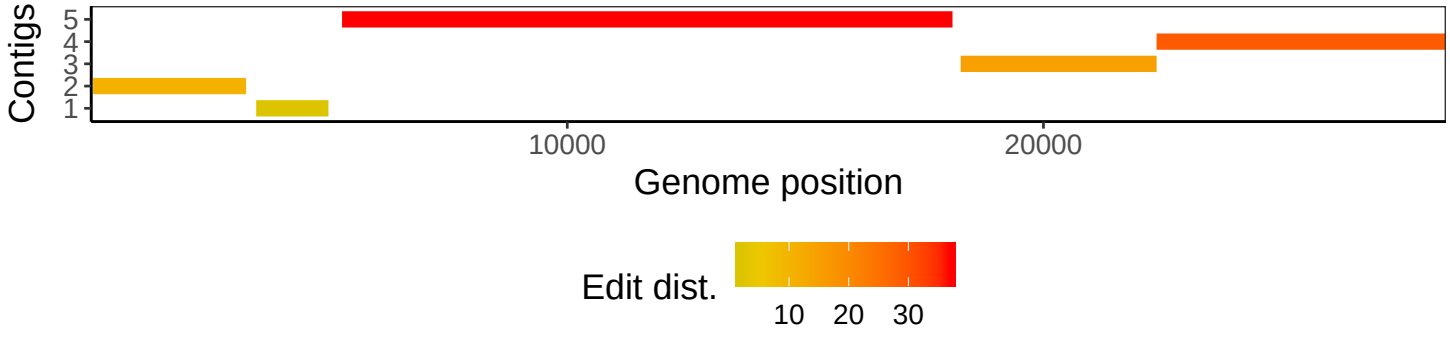
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.3.3
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	3.0.0
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