

COVID-19 subject UPHS-0008

2021-03-25

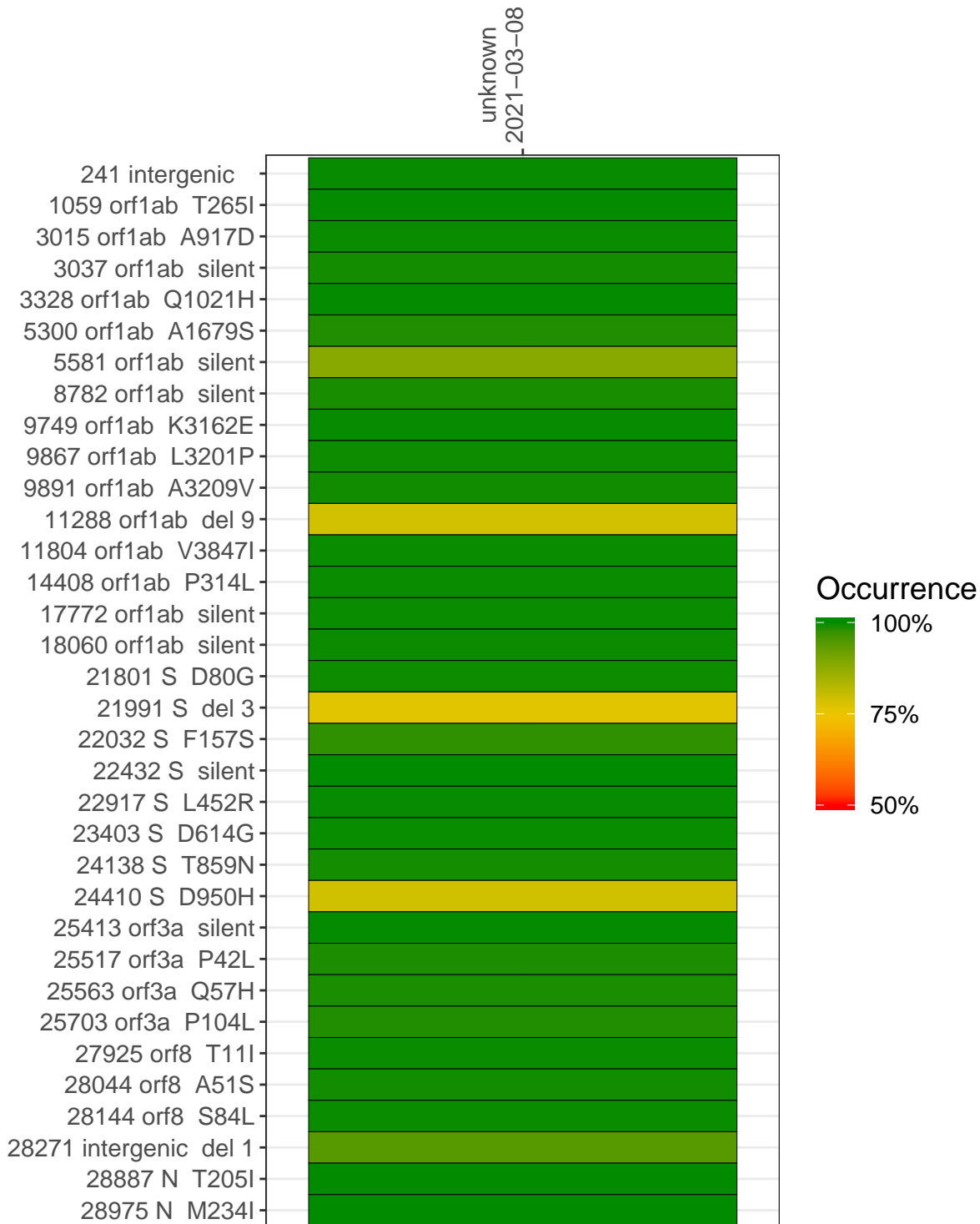
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP0941-1	single experiment	NA	unknown	2021-03-08	29.82	B.1	99.8%	99.8%

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	3394	
1059 orf1ab T265I	5880	
3015 orf1ab A917D	8985	
3037 orf1ab silent	7140	
3328 orf1ab Q1021H	9461	
5300 orf1ab A1679S	3035	
5581 orf1ab silent	7743	
8782 orf1ab silent	11202	
9749 orf1ab K3162E	4726	
9867 orf1ab L3201P	982	
9891 orf1ab A3209V	1568	
11288 orf1ab del 9	14634	
11804 orf1ab V3847I	16471	
14408 orf1ab P314L	13918	
17772 orf1ab silent	5712	
18060 orf1ab silent	8555	
21801 S D80G	7804	
21991 S del 3	1199	
22032 S F157S	1499	
22432 S silent	119	
22917 S L452R	1135	
23403 S D614G	21718	
24138 S T859N	3663	
24410 S D950H	4172	
25413 orf3a silent	7405	
25517 orf3a P42L	6678	
25563 orf3a Q57H	6467	
25703 orf3a P104L	6612	
27925 orf8 T11I	10000	
28044 orf8 A51S	11002	
28144 orf8 S84L	4806	
28271 intergenic del 1	6207	
28887 N T205I	1776	
28975 N M234I	1454	
	VSP0941-1	

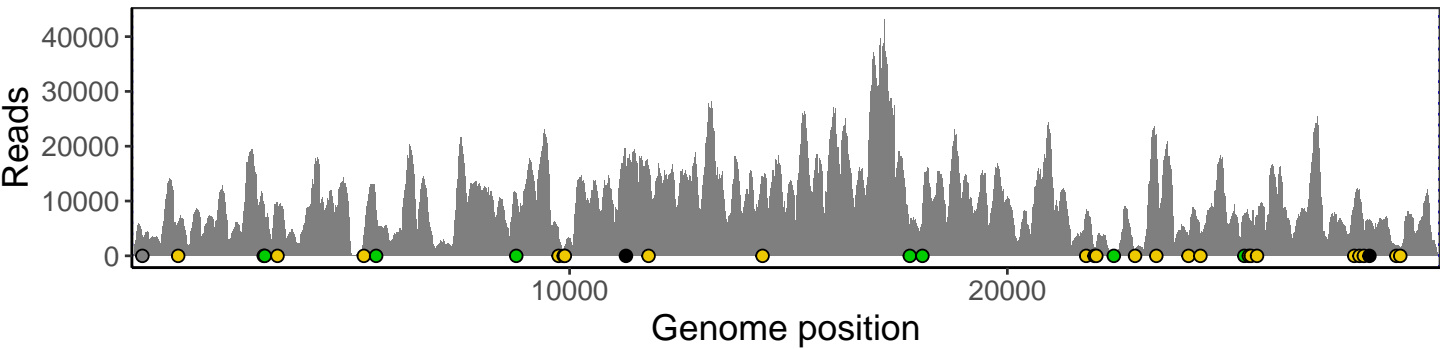
Base change

- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

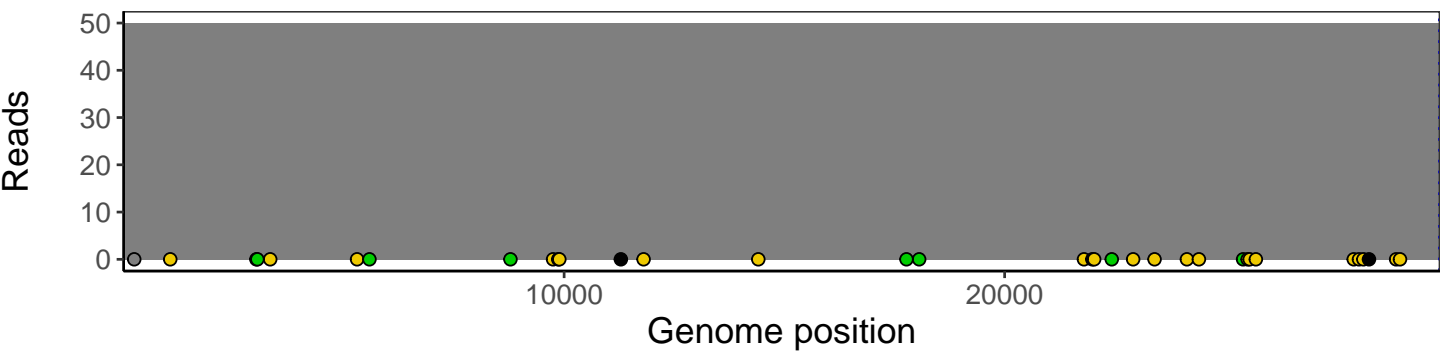
Analyses of individual experiments and composite results

VSP0941-1 | 2021-03-08 | unknown | UPHS-0008 | genomes | single experiment

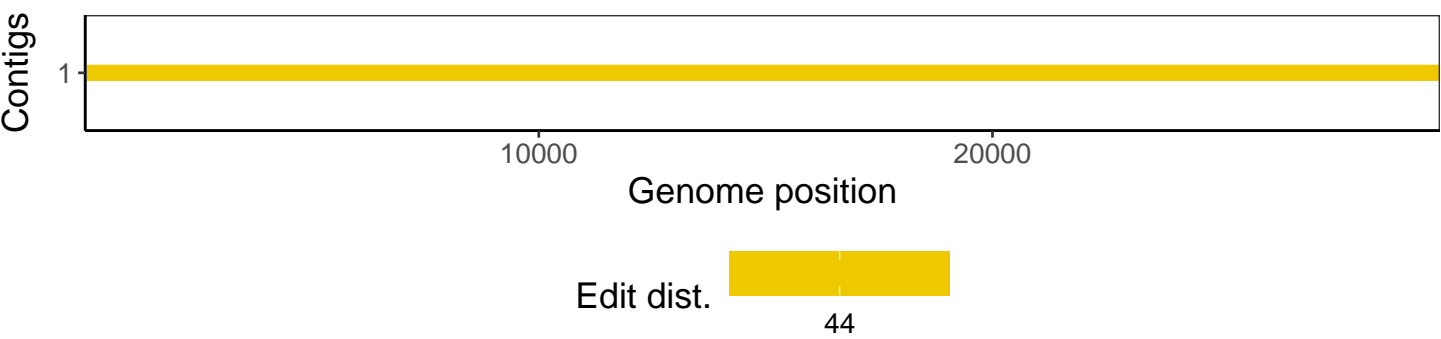
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according to variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	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