

# COVID-19 subject H2102170736

*2021-03-29*

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.

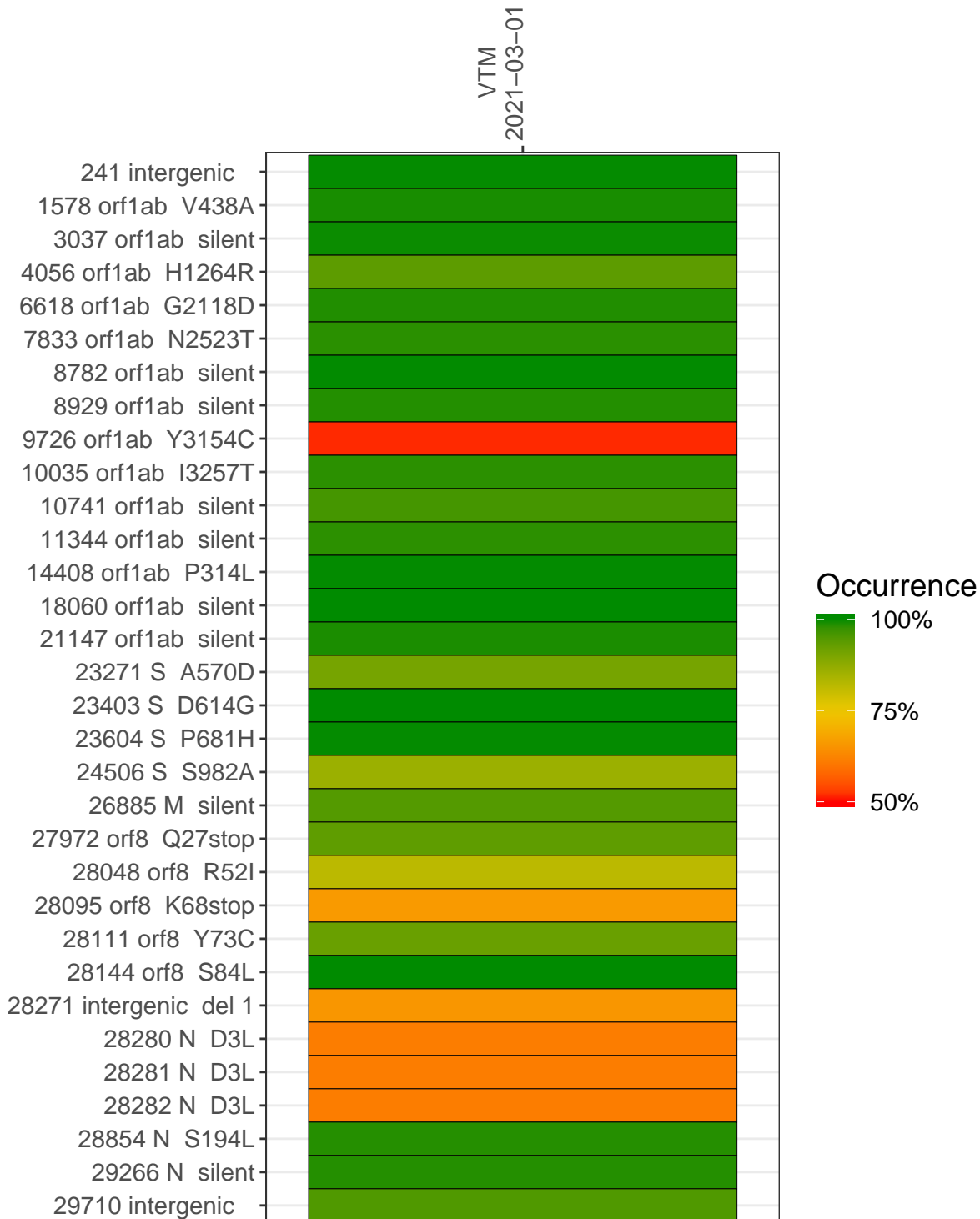
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with > 90% sequence coverage.

Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0677-1	single experiment	NA	VTM	2021-03-01	29.83	B.1	99.8%	99.4%

## 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.



VTM

241 intergenic	2759
1578 orf1ab V438A	3324
3037 orf1ab silent	2369
4056 orf1ab H1264R	1519
6618 orf1ab G2118D	4782
7833 orf1ab N2523T	5355
8782 orf1ab silent	4860
8929 orf1ab silent	5722
9726 orf1ab Y3154C	5595
10035 orf1ab I3257T	388
10741 orf1ab silent	3202
11344 orf1ab silent	6288
14408 orf1ab P314L	2629
18060 orf1ab silent	7155
21147 orf1ab silent	3486
23271 S A570D	64
23403 S D614G	52
23604 S P681H	12615
24506 S S982A	57
26885 M silent	1002
27972 orf8 Q27stop	75
28048 orf8 R52I	66
28095 orf8 K68stop	62
28111 orf8 Y73C	64
28144 orf8 S84L	50
28271 intergenic del 1	81
28280 N D3L	67
28281 N D3L	67
28282 N D3L	67
28854 N S194L	368
29266 N silent	3627
29710 intergenic	1537

Base change

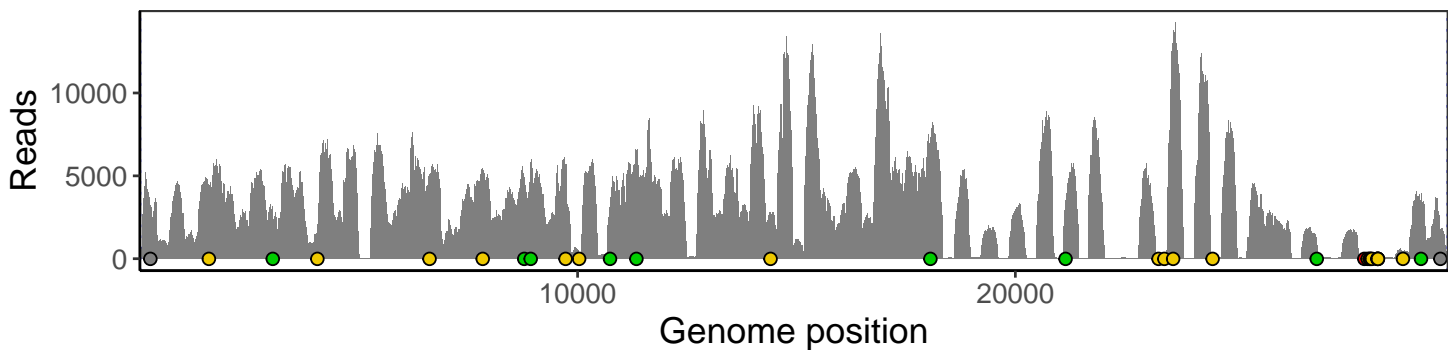


VSP0677-1

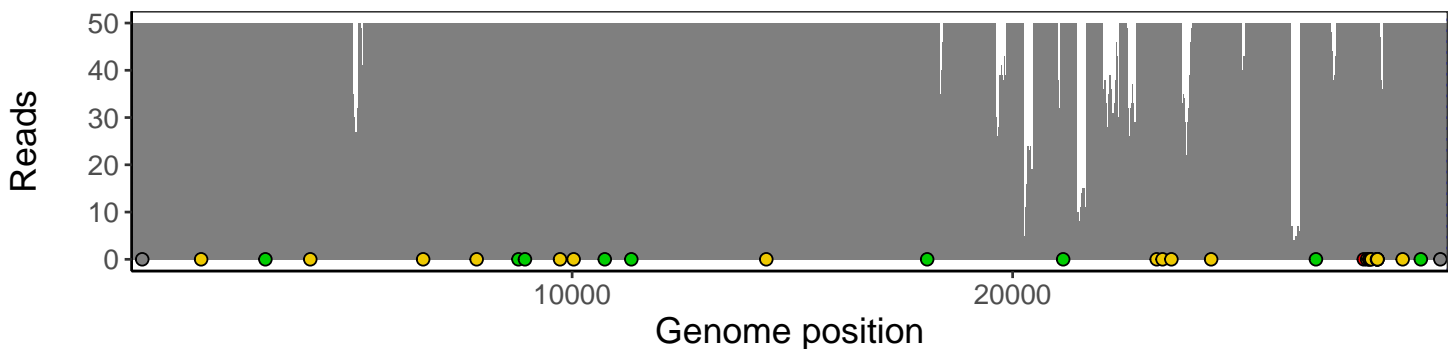
## Analyses of individual experiments and composite results

VSP0677-1 | 2021-03-01 | VTM | H2102170736 | genomes | single experiment

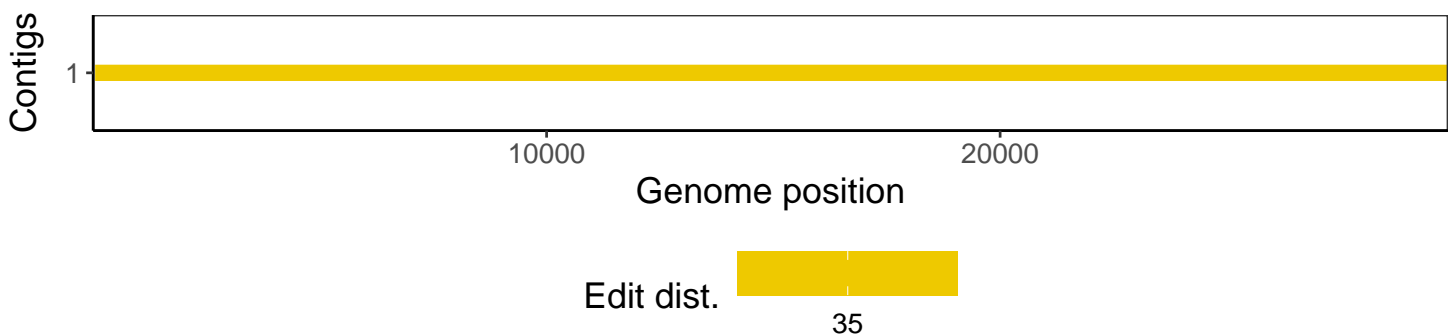
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according to variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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