

# COVID-19 subject H2101300048

*2021-03-01*

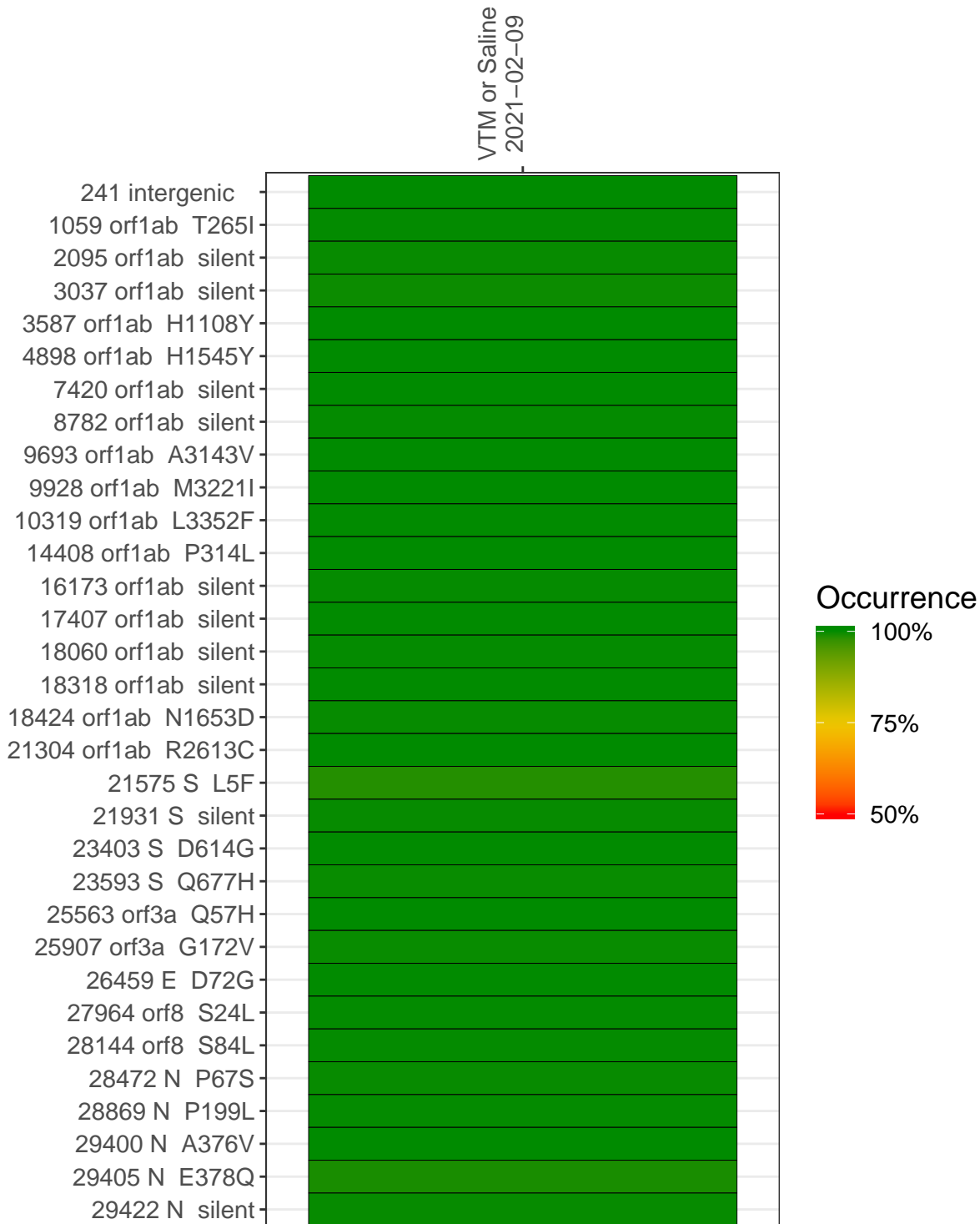
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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 ( $\geq 5$ reads)
VSP0660-1	single experiment	NA	VTM or Saline	2021-02-09	29.83	B.1.2	99.8%	99.7%

## 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 or Saline 2021-02-09	
241 intergenic	9702	
1059 orf1ab T265I	4882	
2095 orf1ab silent	10142	
3037 orf1ab silent	1984	
3587 orf1ab H1108Y	2412	
4898 orf1ab H1545Y	4237	
7420 orf1ab silent	3433	
8782 orf1ab silent	2912	
9693 orf1ab A3143V	491	
9928 orf1ab M3221I	1528	
10319 orf1ab L3352F	6457	
14408 orf1ab P314L	4590	
16173 orf1ab silent	1575	
17407 orf1ab silent	24388	
18060 orf1ab silent	1839	
18318 orf1ab silent	6319	
18424 orf1ab N1653D	10855	
21304 orf1ab R2613C	898	
21575 S L5F	148	
21931 S silent	2916	
23403 S D614G	35903	
23593 S Q677H	6917	
25563 orf3a Q57H	7620	
25907 orf3a G172V	4731	
26459 E D72G	5738	
27964 orf8 S24L	7053	
28144 orf8 S84L	14688	
28472 N P67S	18277	
28869 N P199L	3225	
29400 N A376V	1151	
29405 N E378Q	1181	
29422 N silent	1403	
	VSP0660-1	

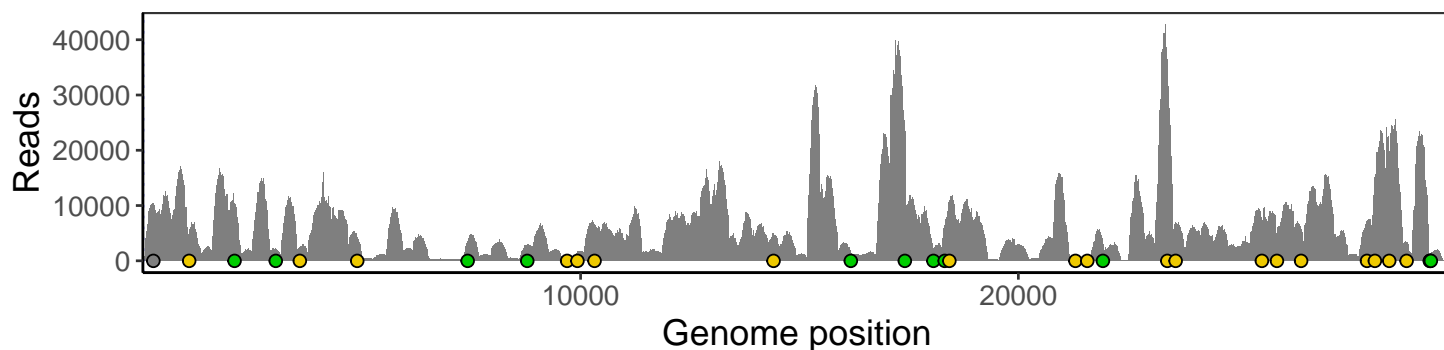
Base change

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

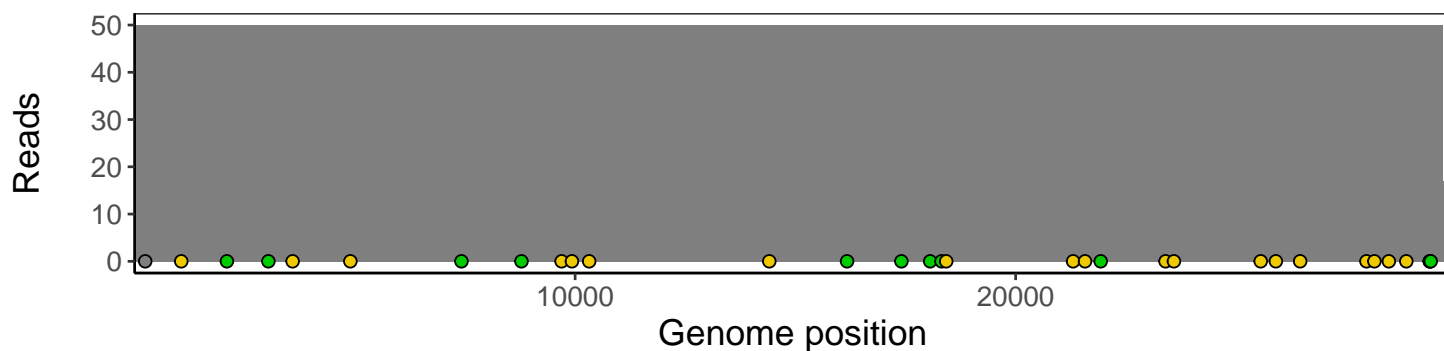
## Analyses of individual experiments and composite results

VSP0660-1 | 2021-02-09 | VTM or Saline | H2101300048 | genomes | single experiment

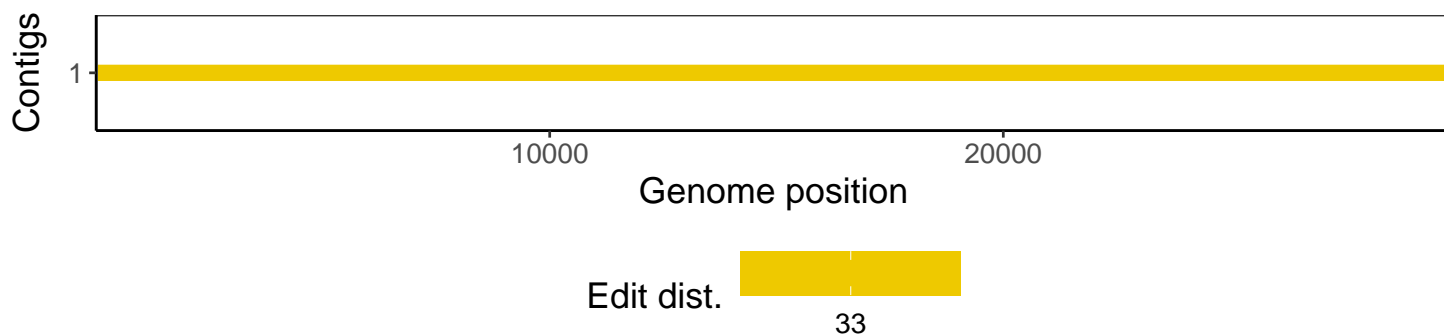
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according to variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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