

# COVID-19 subject H2102030297

*2021-05-05*

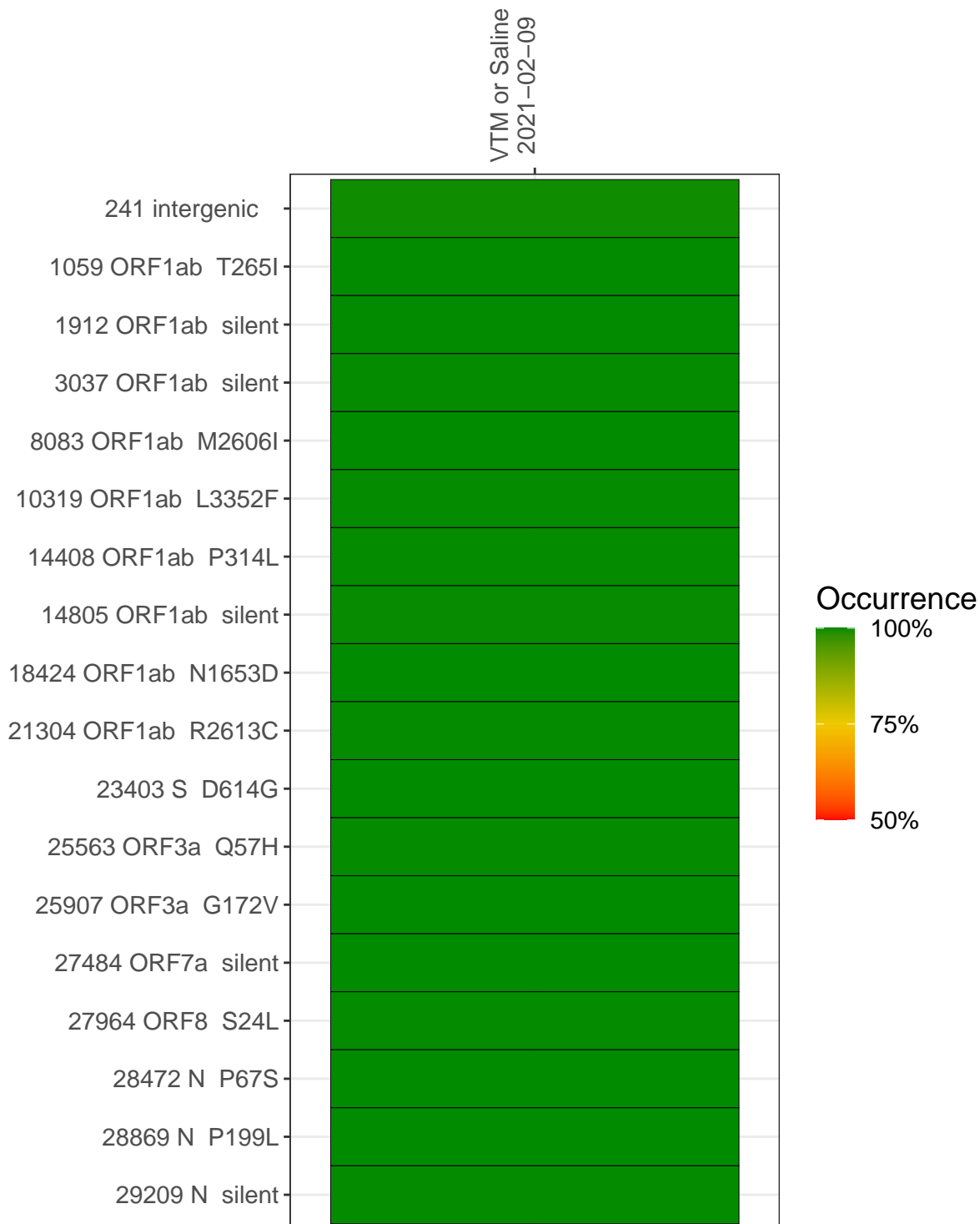
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP0671	composite	NA	VTM or Saline	2021-02-09	29.82	B.1.2	99.9%	99.7%
VSP0671-1	single experiment	NA	VTM or Saline	2021-02-09	NA	NA	NA	NA
VSP0671-2	single experiment	NA	VTM or Saline	2021-02-09	29.82	B.1.2	99.9%	99.7%

## Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/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.



VTM or Saline  
2021-02-09

241 intergenic	3263
1059 ORF1ab T265I	4933
1912 ORF1ab silent	10436
3037 ORF1ab silent	3257
8083 ORF1ab M2606I	5626
10319 ORF1ab L3352F	4975
14408 ORF1ab P314L	3922
14805 ORF1ab silent	6830
18424 ORF1ab N1653D	5429
21304 ORF1ab R2613C	1690
23403 S D614G	7606
25563 ORF3a Q57H	4717
25907 ORF3a G172V	5208
27484 ORF7a silent	6301
27964 ORF8 S24L	3855
28472 N P67S	5538
28869 N P199L	1601
29209 N silent	3351

Base change

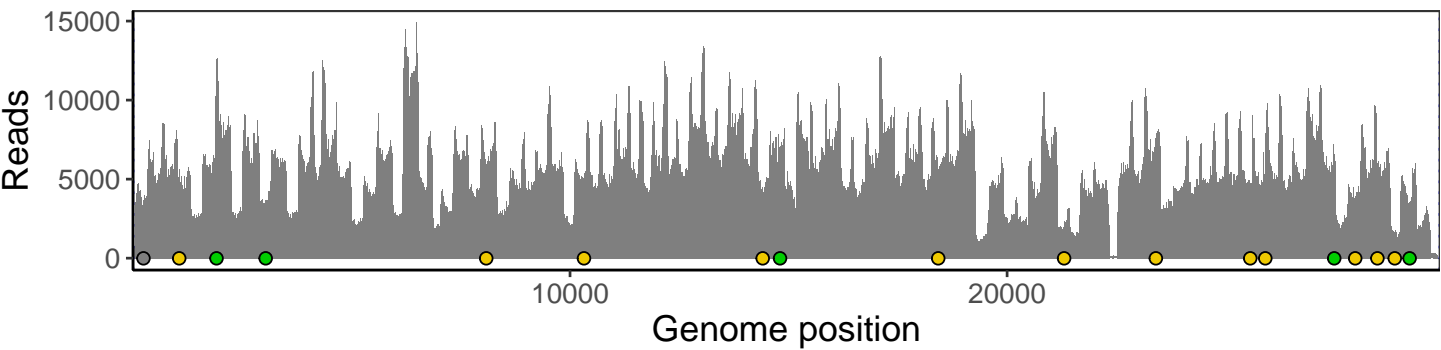
- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

VSP0671-2

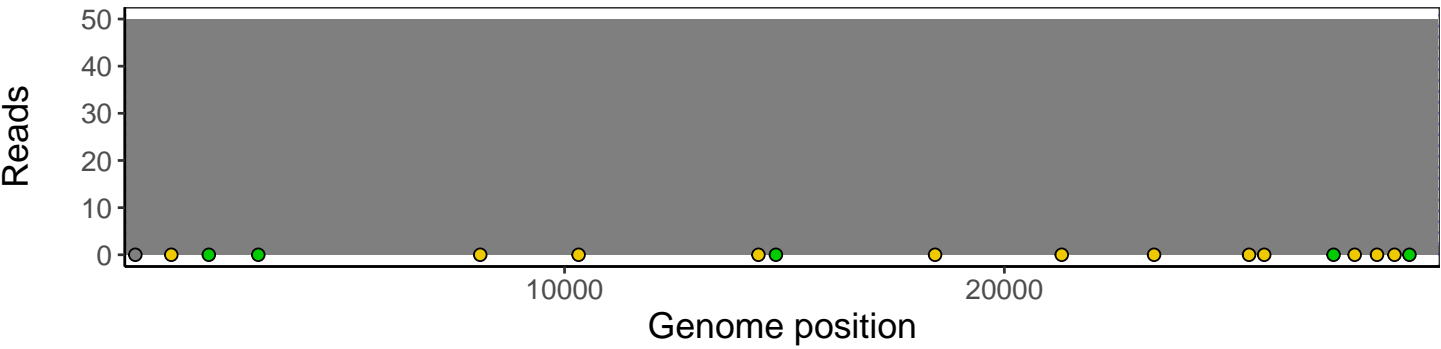
# Analyses of individual experiments and composite results

VSP0671 | 2021-02-09 | VTM or Saline | H2102030297 | composite result

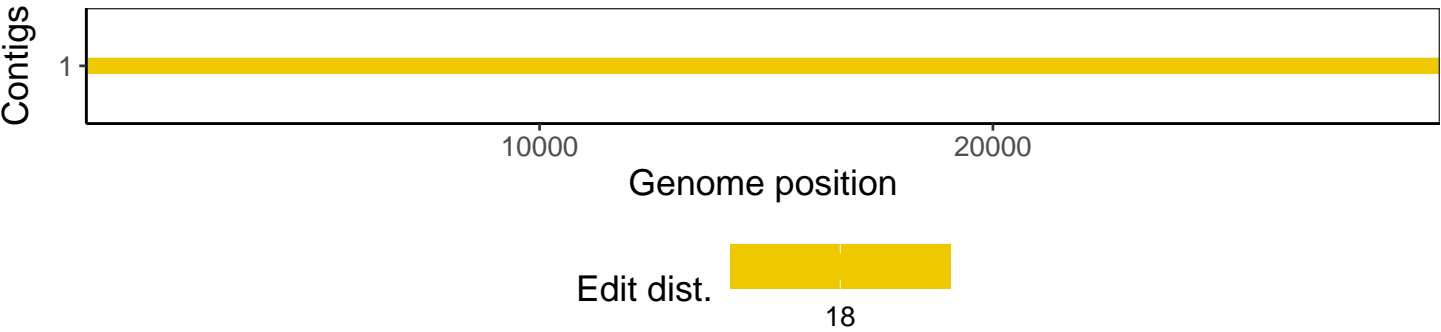
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according to variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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.

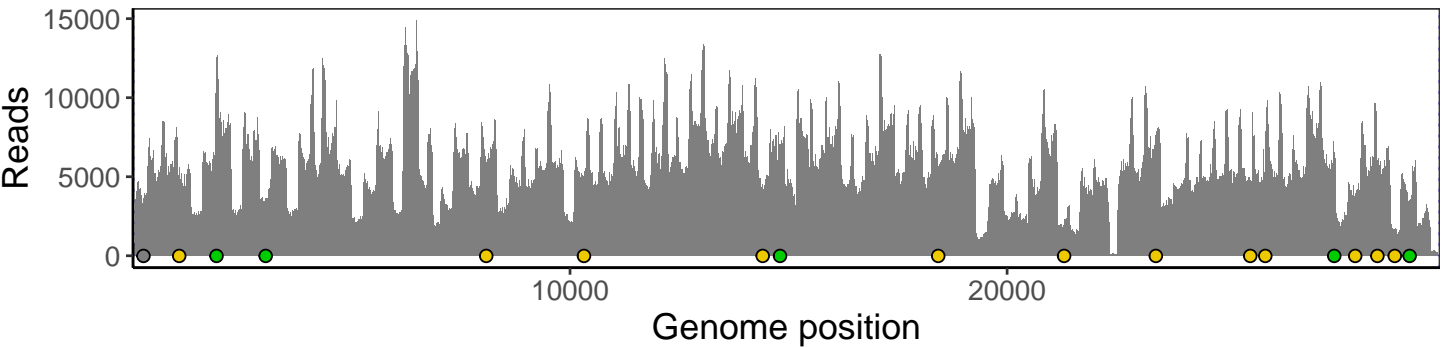


VSP0671-1 | 2021-02-09 | VTM or Saline | H2102030297 | genomes | single experiment

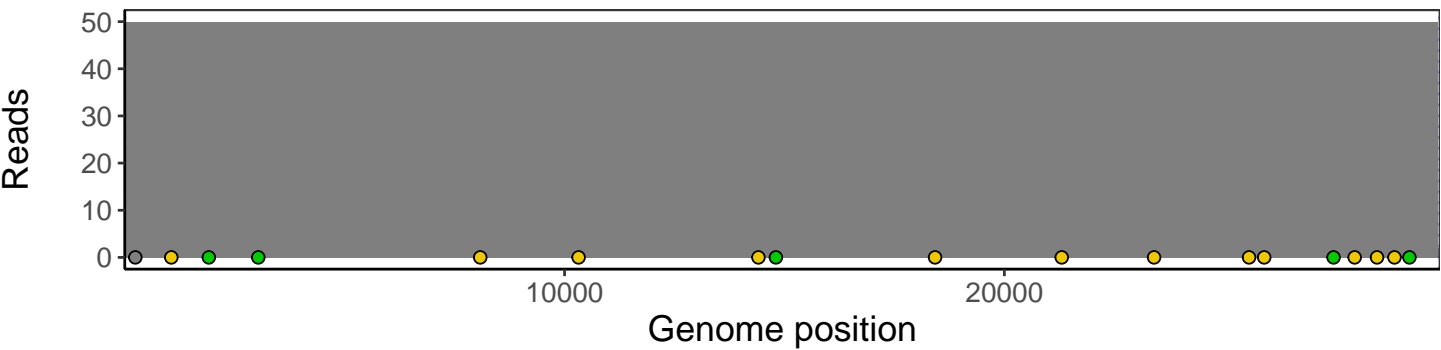
No pileup data available.

No contig data available.

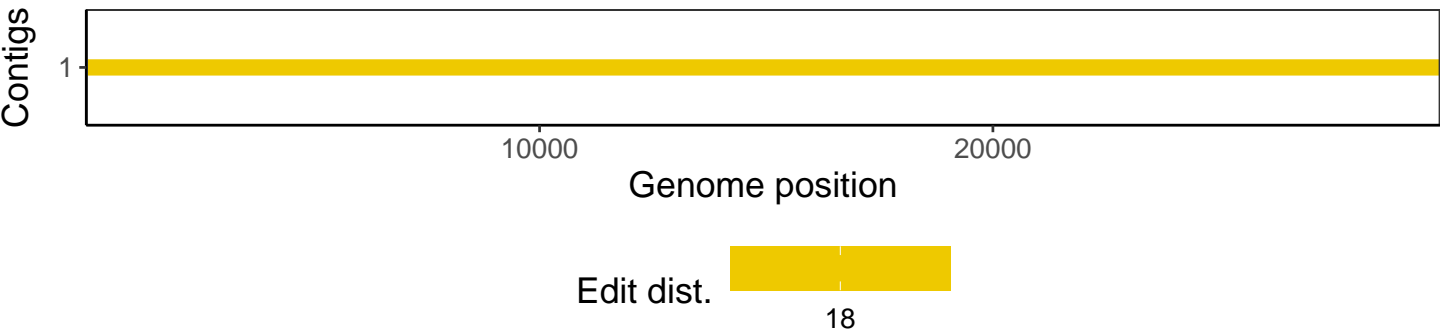
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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 htlib 1.10
bcftools	1.10.2-34-g1a12af0-dirty Using htlib 1.10.2-57-gf58a6f3
pangolin	2.3.8
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