

COVID-19 subject H2102040793

2021-05-05

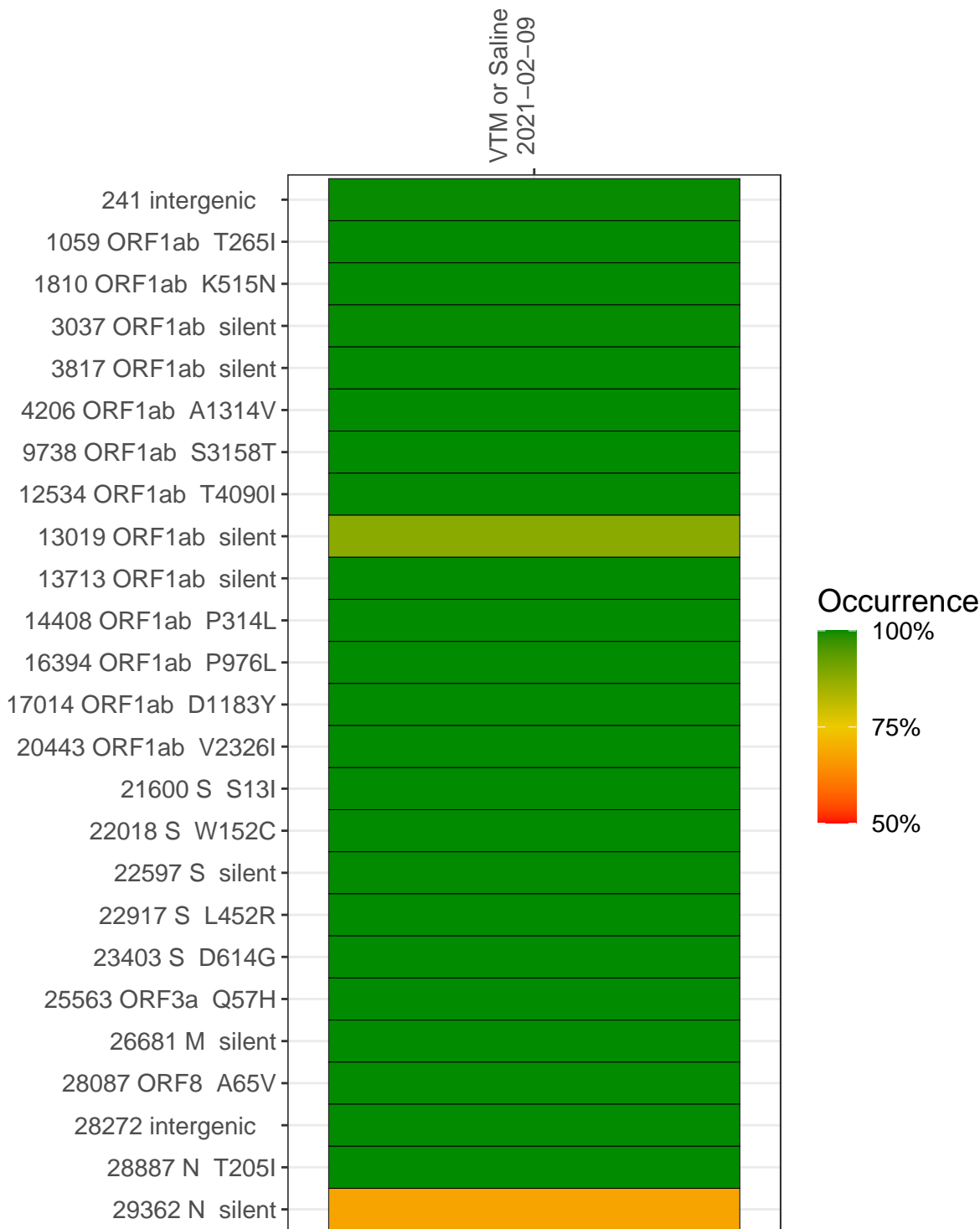
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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 (≥ 5 reads)
VSP0663-1	single experiment	NA	VTM or Saline	2021-02-09	29.85	B.1.427	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	6689
1059 ORF1ab T265I	2587
1810 ORF1ab K515N	6557
3037 ORF1ab silent	1837
3817 ORF1ab silent	4300
4206 ORF1ab A1314V	8369
9738 ORF1ab S3158T	7675
12534 ORF1ab T4090I	7057
13019 ORF1ab silent	4446
13713 ORF1ab silent	5078
14408 ORF1ab P314L	5146
16394 ORF1ab P976L	4677
17014 ORF1ab D1183Y	14810
20443 ORF1ab V2326I	1895
21600 S S13I	1373
22018 S W152C	319
22597 S silent	9933
22917 S L452R	4265
23403 S D614G	27344
25563 ORF3a Q57H	5628
26681 M silent	10976
28087 ORF8 A65V	7554
28272 intergenic	6963
28887 N T205I	2310
29362 N silent	2909

Base change

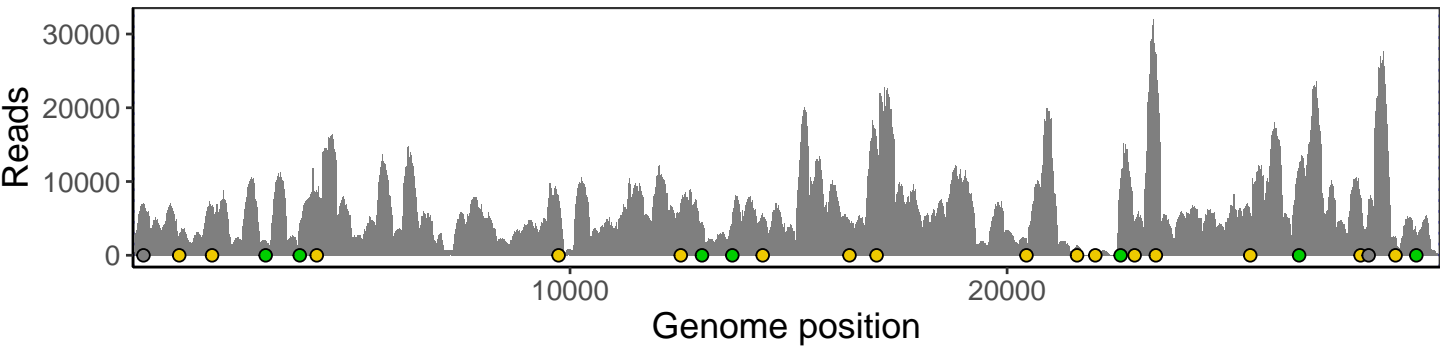
- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

VSP0663-1

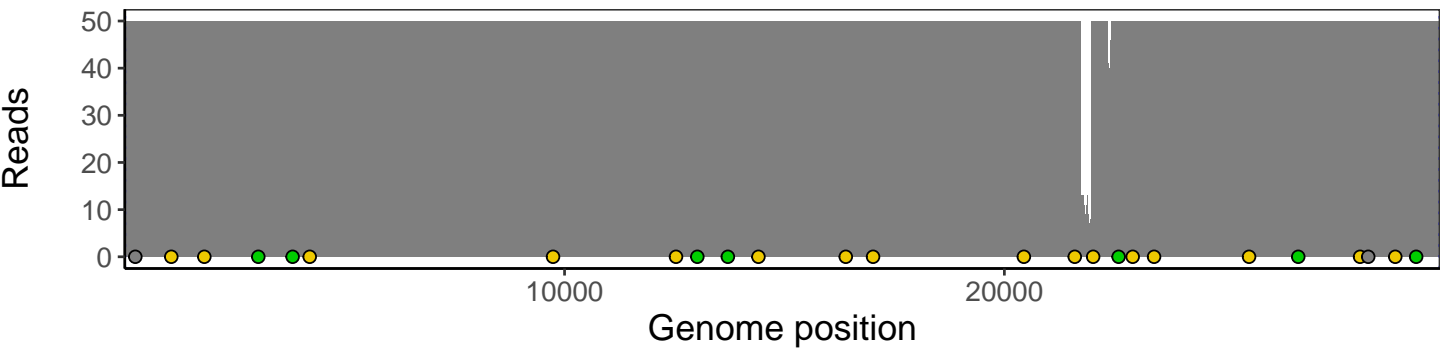
Analyses of individual experiments and composite results

VSP0663-1 | 2021-02-09 | VTM or Saline | H2102040793 | genomes | single experiment

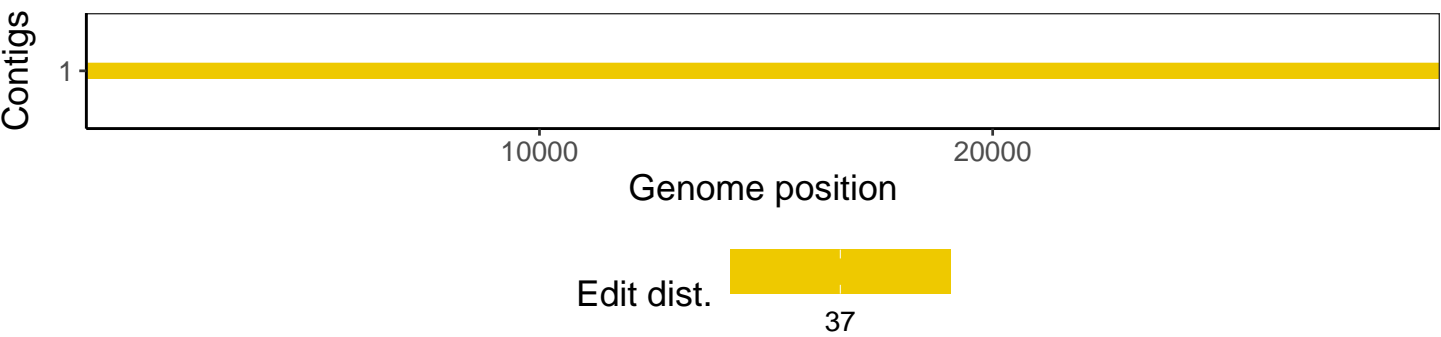
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according to variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.3.8
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