

COVID-19 subject H2102030405

2021-03-01

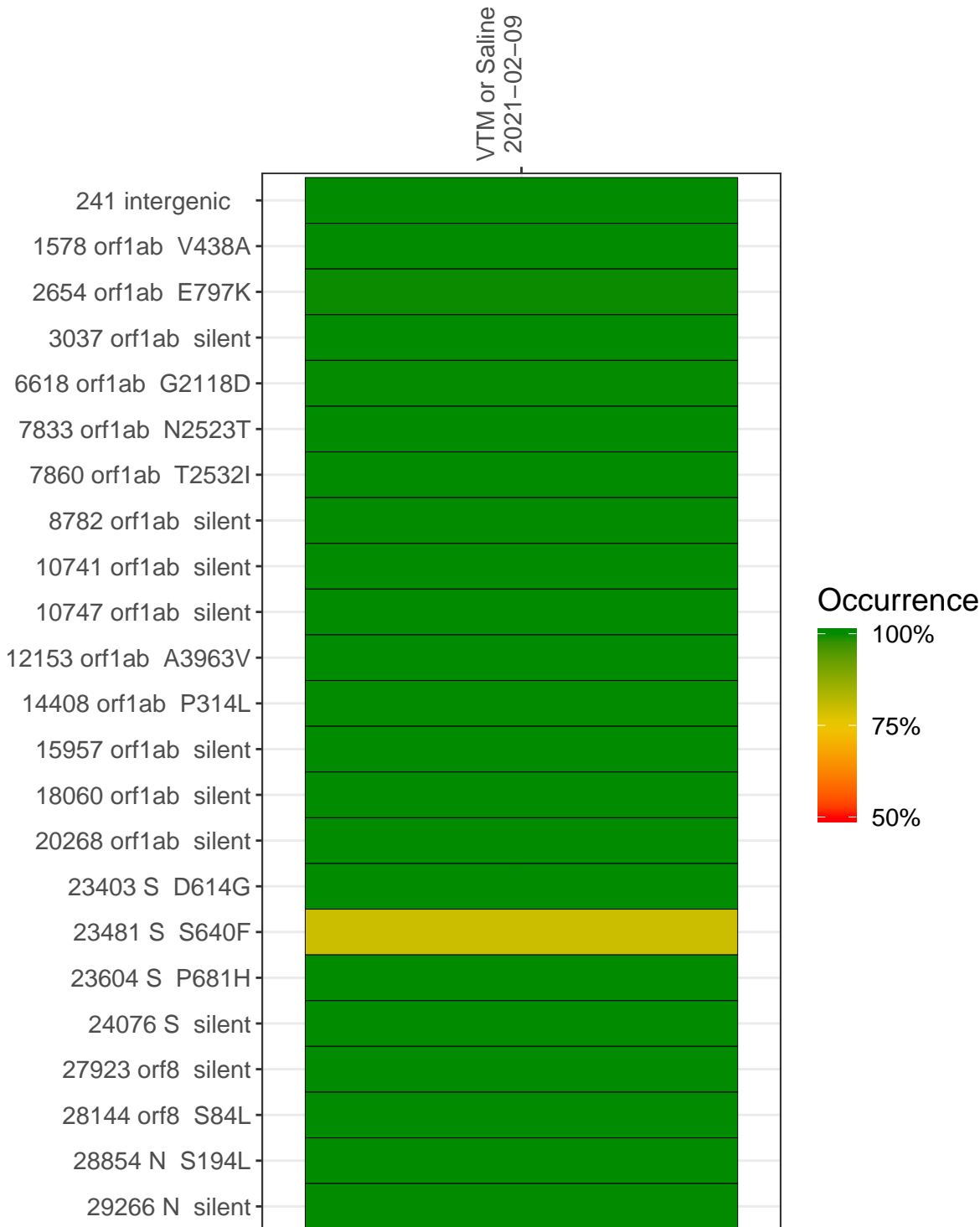
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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 (≥ 5 reads)
VSP0670-1	single experiment	NA	VTM or Saline	2021-02-09	12.39	B.1.243	92.6%	91.0%

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	4030
1578 orf1ab V438A	626
2654 orf1ab E797K	2248
3037 orf1ab silent	824
6618 orf1ab G2118D	1634
7833 orf1ab N2523T	2440
7860 orf1ab T2532I	2220
8782 orf1ab silent	2222
10741 orf1ab silent	4719
10747 orf1ab silent	4779
12153 orf1ab A3963V	7320
14408 orf1ab P314L	6527
15957 orf1ab silent	6087
18060 orf1ab silent	4532
20268 orf1ab silent	1983
23403 S D614G	9295
23481 S S640F	7823
23604 S P681H	3732
24076 S silent	6069
27923 orf8 silent	3652
28144 orf8 S84L	4429
28854 N S194L	900
29266 N silent	2801

Base change

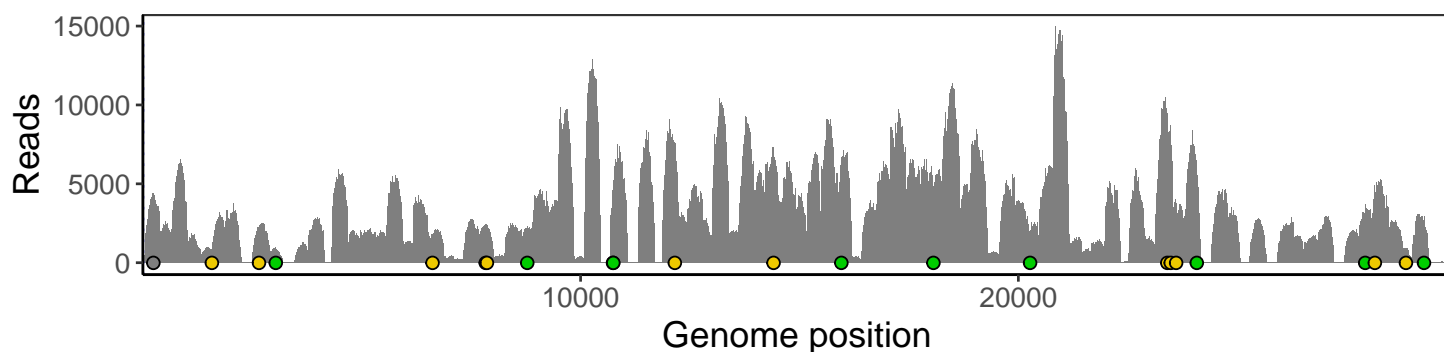


VSP0670-1

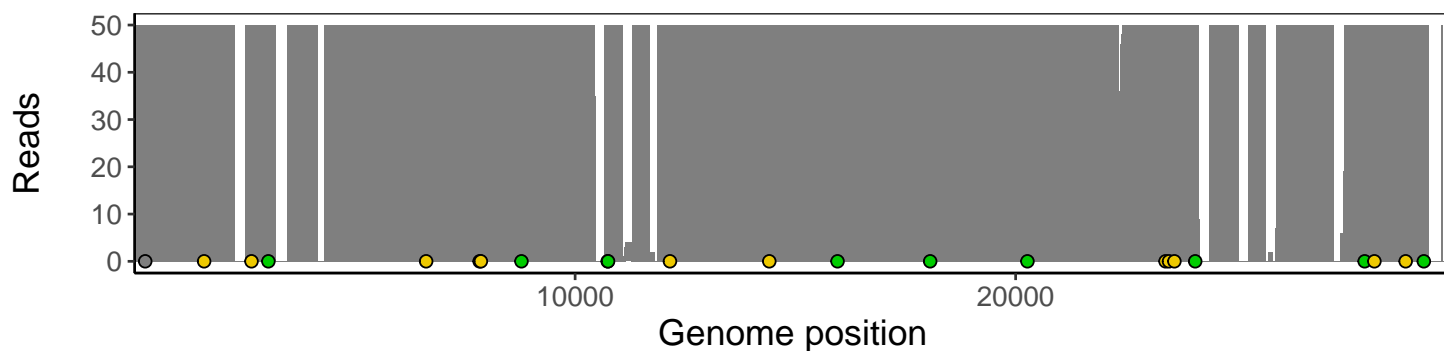
Analyses of individual experiments and composite results

VSP0670-1 | 2021-02-09 | VTM or Saline | H2102030405 | genomes | single experiment

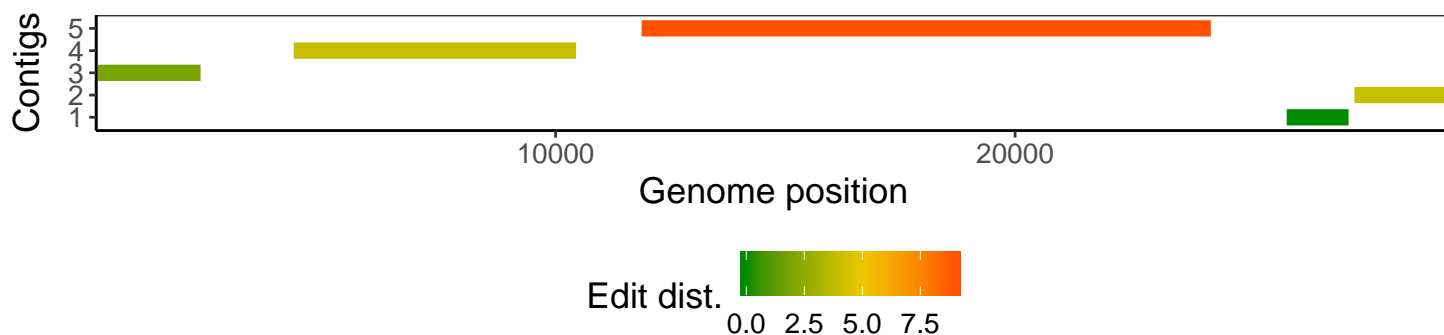
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.3.3
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