

COVID-19 subject H2102230845

2021-03-29

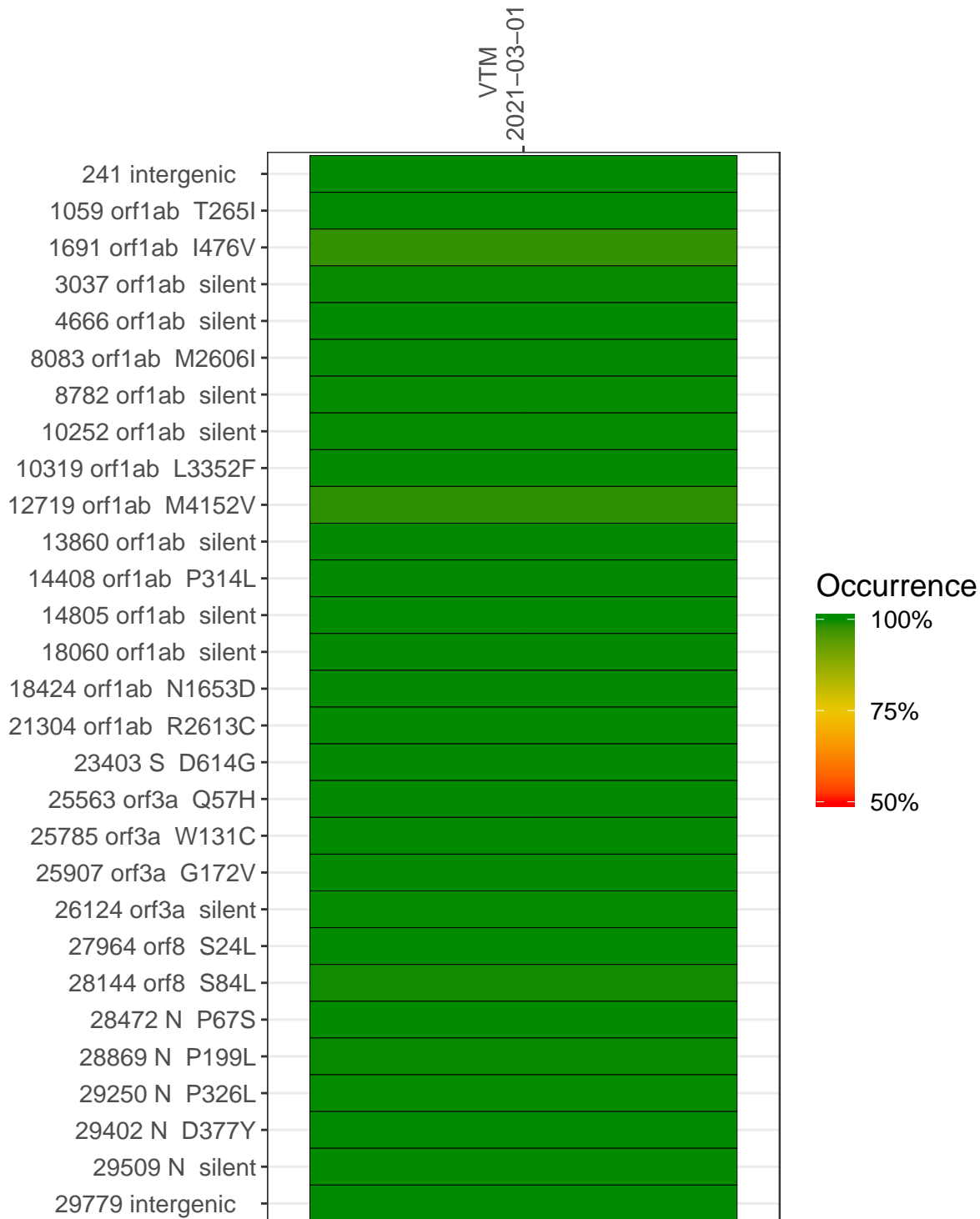
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP0682-1	single experiment	NA	VTM	2021-03-01	29.82	B.1.2	99.8%	99.8%

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	3713
1059 orf1ab T265I	5915
1691 orf1ab I476V	9748
3037 orf1ab silent	6705
4666 orf1ab silent	8850
8083 orf1ab M2606I	10402
8782 orf1ab silent	10318
10252 orf1ab silent	14160
10319 orf1ab L3352F	14558
12719 orf1ab M4152V	15719
13860 orf1ab silent	13708
14408 orf1ab P314L	15670
14805 orf1ab silent	9154
18060 orf1ab silent	6125
18424 orf1ab N1653D	14987
21304 orf1ab R2613C	6421
23403 S D614G	12718
25563 orf3a Q57H	8380
25785 orf3a W131C	7101
25907 orf3a G172V	5012
26124 orf3a silent	8665
27964 orf8 S24L	22472
28144 orf8 S84L	13434
28472 N P67S	11892
28869 N P199L	1999
29250 N P326L	11591
29402 N D377Y	10077
29509 N silent	13337
29779 intergenic	391

Base change

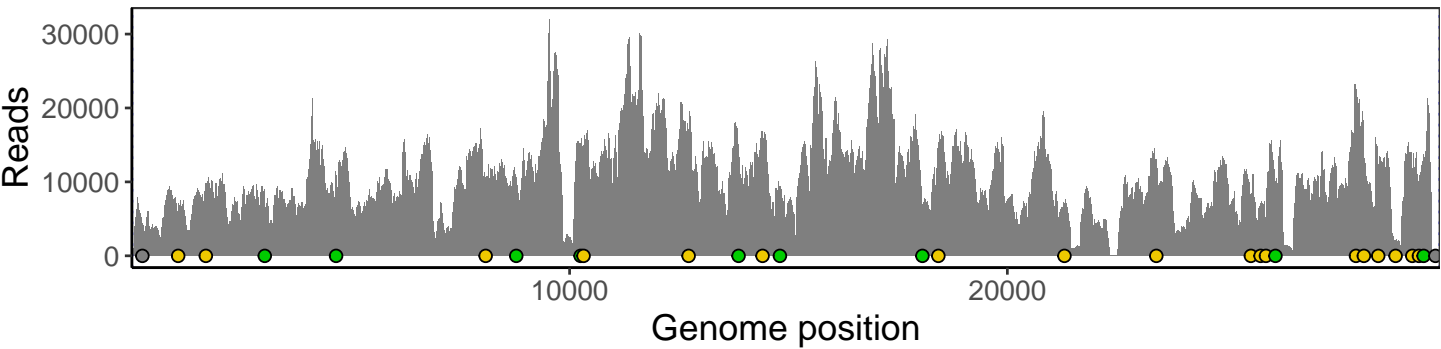
Expected
A
T
C
G
N
Ins/Del
No data

VSP0682-1

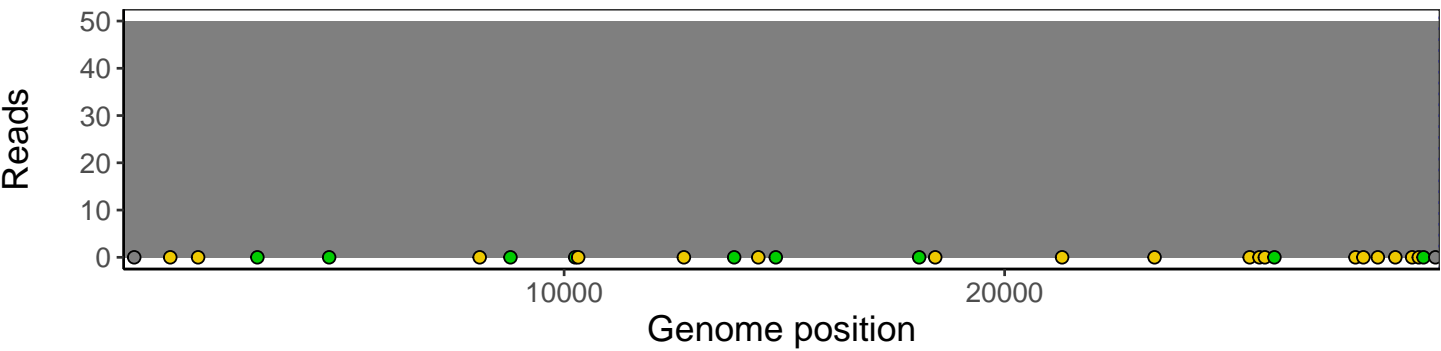
Analyses of individual experiments and composite results

VSP0682-1 | 2021-03-01 | VTM | H2102230845 | genomes | single experiment

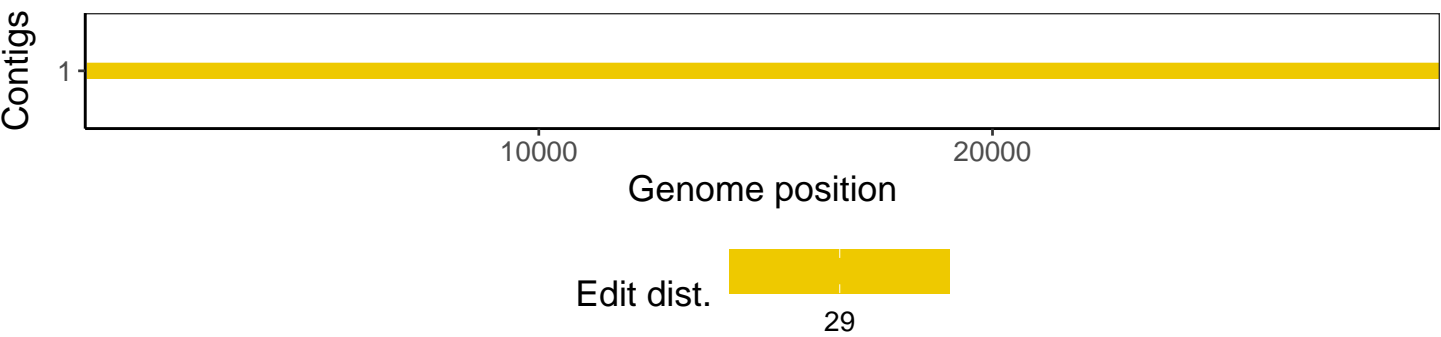
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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