COVID-19 subject UPHS-0138

2021-04-17

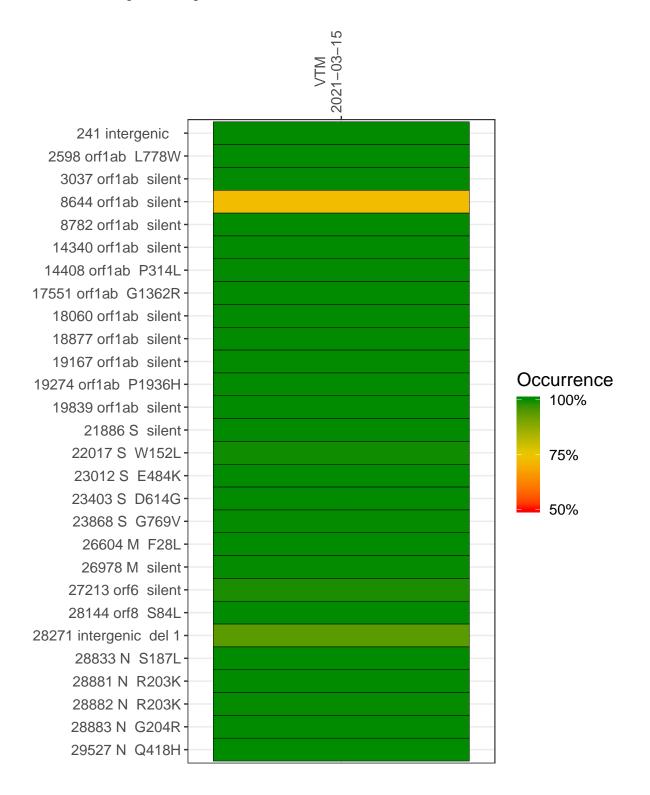
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with > 90% sequence coverage.

Table 1. Sample summary.

Experiment	Туре	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1123-1	single experiment	NA	VTM	2021-03-15	29.90	R.1	99.8%	99.8%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/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 2021-03-15

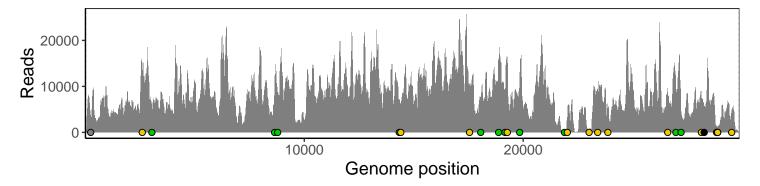
	2021-00-10
241 intergenic	3342
2598 orf1ab L778W	12309
3037 orf1ab silent	4923
8644 orf1ab silent	11387
8782 orf1ab silent	7929
14340 orf1ab silent	6641
14408 orf1ab P314L	5480
17551 orf1ab G1362R	9725
18060 orf1ab silent	7409
18877 orf1ab silent	15187
19167 orf1ab silent	8889
19274 orf1ab P1936H	14636
19839 orf1ab silent	11347
21886 S silent	71
22017 S W152L	3146
23012 S E484K	476
23403 S D614G	9402
23868 S G769V	6189
26604 M F28L	8958
26978 M silent	6916
27213 orf6 silent	4415
28144 orf8 S84L	7564
28271 intergenic del 1	5330
28833 N S187L	1362
28881 N R203K	878
28882 N R203K	875
28883 N G204R	877
29527 N Q418H	3900
	<u></u>
	123
	VSP1123-1
	>



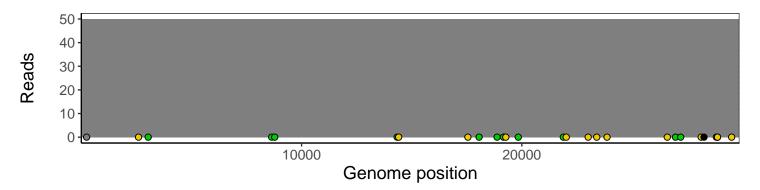
Analyses of individual experiments and composite results

VSP1123-1 | 2021-03-15 | VTM | UPHS-0138 | genomes | single experiment

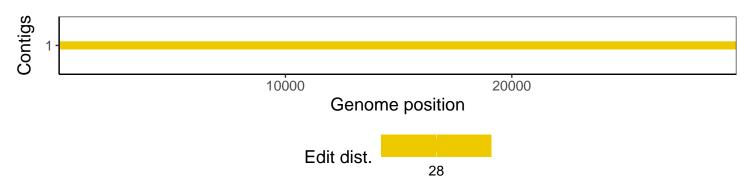
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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 htslib 1.10
bcftools	1.10.2-34-g1a12af0-dirty Using htslib 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
${\it Genomic Alignments}$	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
$\operatorname{GenomeInfoDb}$	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