COVID-19 subject UPHS-0269

2021-04-17

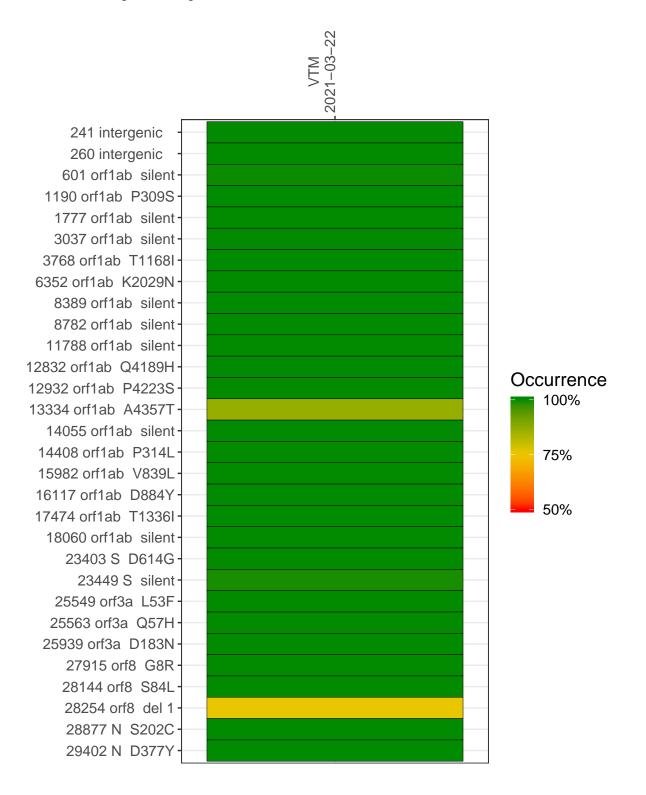
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1314-1	single experiment	NA	VTM	2021-03-22	29.85	B.1.110.3	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 2021-03-22

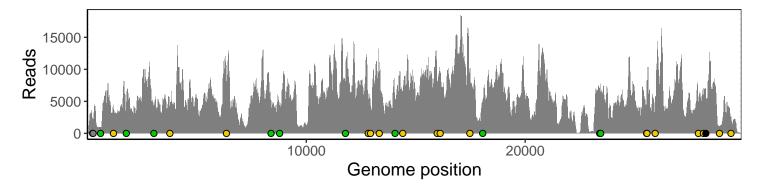
	2021-03-22
241 intergenic	2092
260 intergenic	2080
601 orf1ab silent	926
1190 orf1ab P309S	3727
1777 orf1ab silent	4011
3037 orf1ab silent	3723
3768 orf1ab T1168I	2896
6352 orf1ab K2029N	8494
8389 orf1ab silent	4892
8782 orf1ab silent	4184
11788 orf1ab silent	7760
12832 orf1ab Q4189H	5250
12932 orf1ab P4223S	4299
13334 orf1ab A4357T	10674
14055 orf1ab silent	5179
14408 orf1ab P314L	575 9
15982 orf1ab V839L	8048
16117 orf1ab D884Y	9647
17474 orf1ab T1336I	10012
18060 orf1ab silent	4641
23403 S D614G	6961
23449 S silent	7476
25549 orf3a L53F	7692
25563 orf3a Q57H	7992
25939 orf3a D183N	4862
27915 orf8 G8R	4598
28144 orf8 S84L	6083
28254 orf8 del 1	6163
28877 N S202C	1087
29402 N D377Y	2633



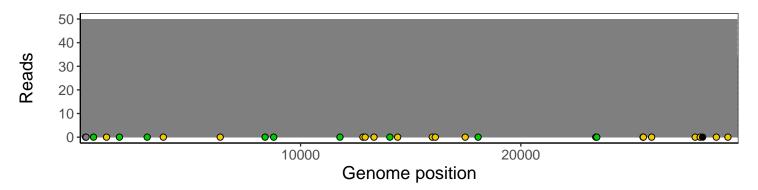
Analyses of individual experiments and composite results

VSP1314-1 | 2021-03-22 | VTM | UPHS-0269 | genomes | single experiment

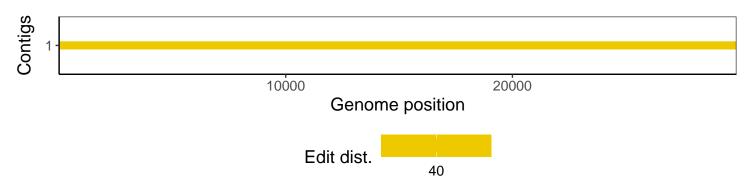
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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