COVID-19 subject UPHS-0534

2021-06-23

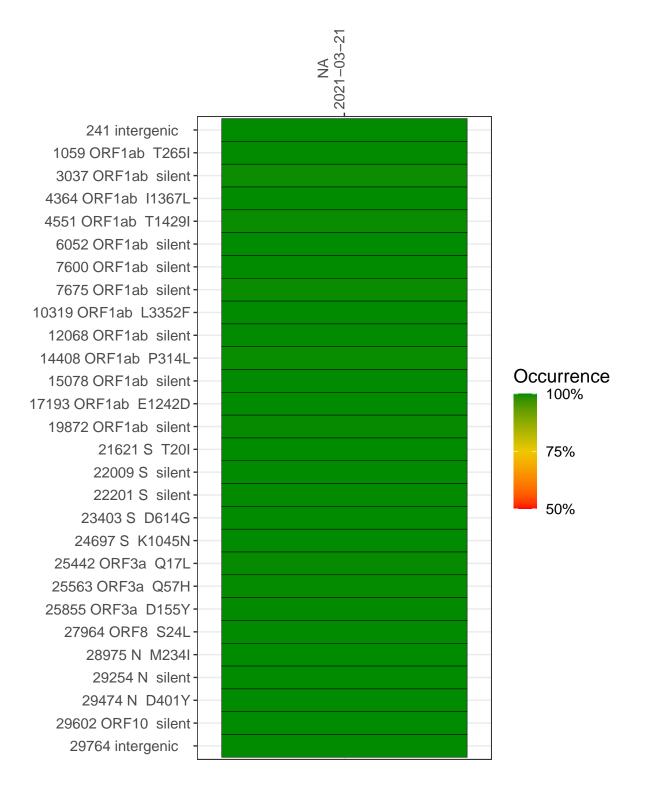
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1660-1	single experiment	NA	NA	2021-03-21	29.82	B.1.588	99.7%	99.6%

Variants shared across samples

The heat map below shows how variants (reference genome /home/common/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.



NA 2021-03-21

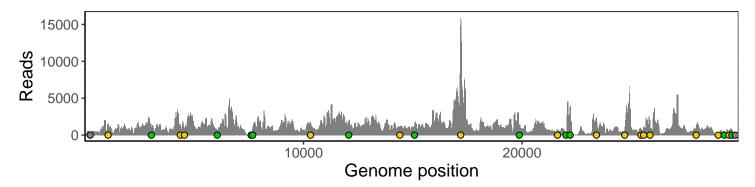
	2021-03-21
241 intergenic	304
1059 ORF1ab T265I	1454
3037 ORF1ab silent	1054
4364 ORF1ab I1367L	1247
4551 ORF1ab T1429I	906
6052 ORF1ab silent	703
7600 ORF1ab silent	16
7675 ORF1ab silent	934
10319 ORF1ab L3352F	1700
12068 ORF1ab silent	1935
14408 ORF1ab P314L	989
15078 ORF1ab silent	1164
17193 ORF1ab E1242D	15893
19872 ORF1ab silent	1303
21621 S T20I	690
22009 S silent	411
22201 S silent	3720
23403 S D614G	2352
24697 S K1045N	616
25442 ORF3a Q17L	665
25563 ORF3a Q57H	1006
25855 ORF3a D155Y	1991
27964 ORF8 S24L	1213
28975 N M234I	225
29254 N silent	548
29474 N D401Y	443
29602 ORF10 silent	780
29764 intergenic	223
	1-0
	999
	VSP1660-1
	*



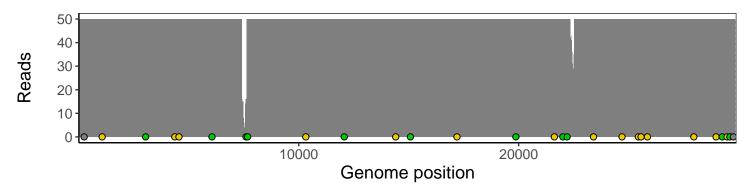
Analyses of individual experiments and composite results

VSP1660-1 | 2021-03-21 | NA | UPHS-0534 | genomes | single experiment

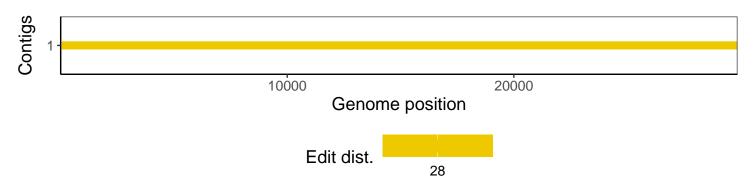
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	3.1.3
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	3.3.3
tidyverse	1.2.1
ShortRead	1.34.2
GenomicAlignments	1.12.2
${\bf Summarized Experiment}$	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