COVID-19 subject UPHS-0648

2021-06-23

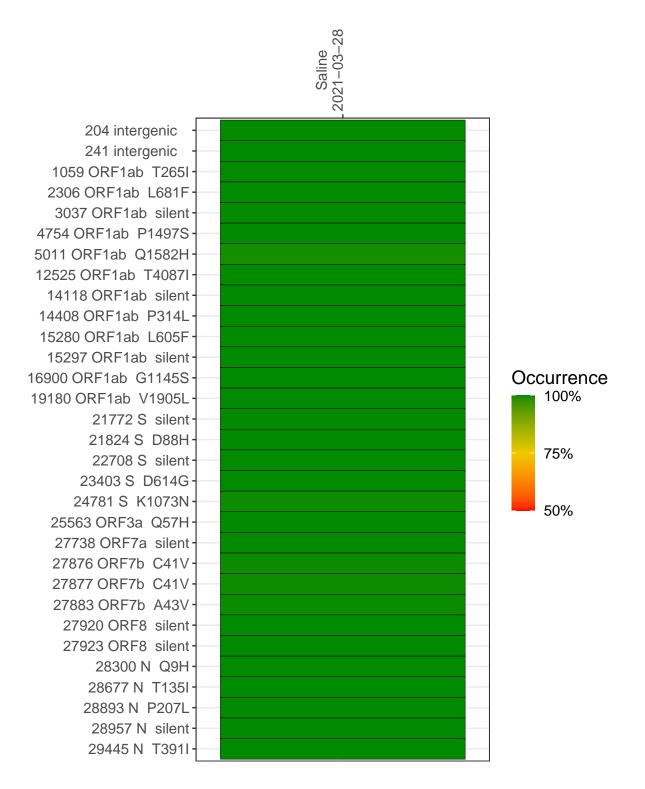
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1833-1	single experiment	NA	Saline	2021-03-28	29.87	B.1.311	99.9%	99.8%

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.



Saline 2021-03-28

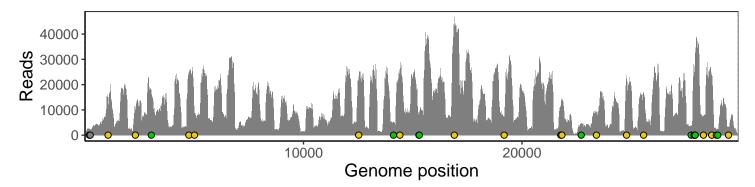
	2021-03-20
204 intergenic	2205
241 intergenic	1962
1059 ORF1ab T265I	12568
2306 ORF1ab L681F	9993
3037 ORF1ab silent	15168
4754 ORF1ab P1497S	21719
5011 ORF1ab Q1582H	7379
12525 ORF1ab T4087I	22230
14118 ORF1ab silent	5029
14408 ORF1ab P314L	23637
15280 ORF1ab L605F	7692
15297 ORF1ab silent	8038
16900 ORF1ab G1145S	38450
19180 ORF1ab V1905L	7139
21772 S silent	11024
21824 S D88H	13748
22708 S silent	4304
23403 S D614G	8981
24781 S K1073N	20356
25563 ORF3a Q57H	15785
27738 ORF7a silent	3033
27876 ORF7b C41V	8616
27877 ORF7b C41V	8616
27883 ORF7b A43V	8955
27920 ORF8 silent	27711
27923 ORF8 silent	26887
28300 N Q9H	6804
28677 N T135I	22491
28893 N P207L	2316
28957 N silent	2214
29445 N T391I	3020
	3 – 7



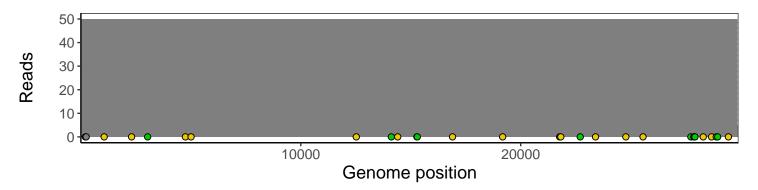
Analyses of individual experiments and composite results

$VSP1833-1 \mid 2021-03-28 \mid Saline \mid UPHS-0648 \mid genomes \mid single \; experiment$

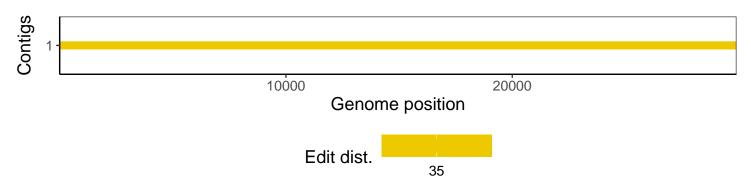
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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