COVID-19 subject UPHS-0813

2021-06-23

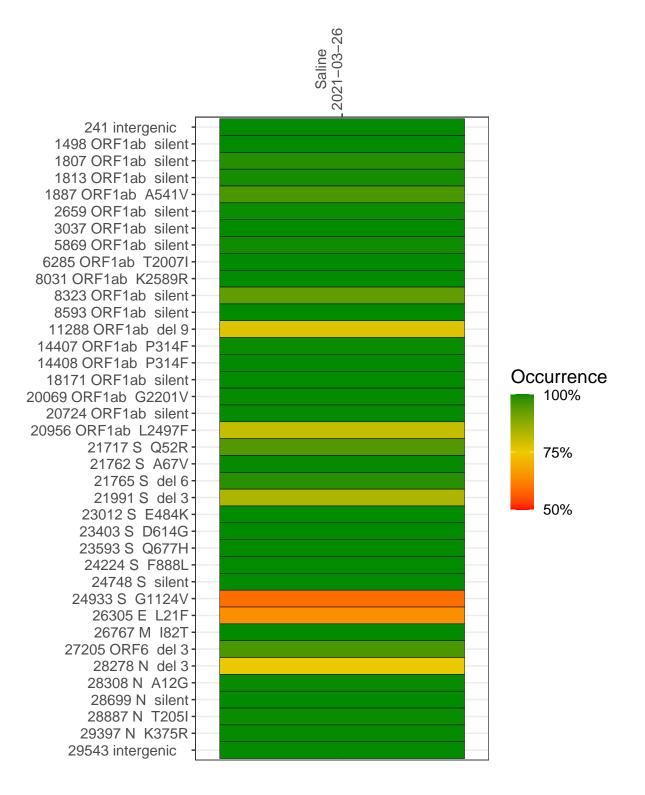
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP2027-2	single experiment	NA	Saline	2021-03-26	29.75	B.1.525	99.6%	99.5%

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 2021-03-26

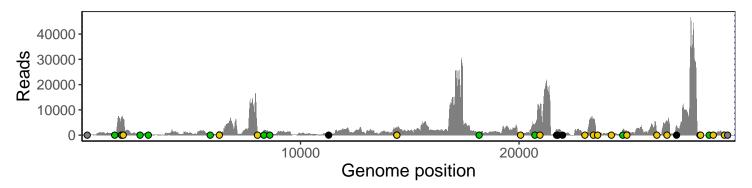
	2021 00 20
241 intergenic	194
1498 ORF1ab silent	1100
1807 ORF1ab silent	6838
1813 ORF1ab silent	6990
1887 ORF1ab A541V	6398
2659 ORF1ab silent	690
3037 ORF1ab silent	490
5869 ORF1ab silent	477
6285 ORF1ab T2007I	1127
8031 ORF1ab K2589R	875
8323 ORF1ab silent	1682
8593 ORF1ab silent	1352
11288 ORF1ab del 9	515
14407 ORF1ab P314F	2413
14408 ORF1ab P314F	2444
18171 ORF1ab silent	1184
20069 ORF1ab G2201V	758
20724 ORF1ab silent	4357
20956 ORF1ab L2497F	11216
21717 S Q52R	2245
21762 S A67V	1278
21765 S del 6	1224
21991 S del 3	899
23012 S E484K	29
23403 S D614G	6213
23593 S Q677H	1394
24224 S F888L	652
24748 S silent	2183
24933 S G1124V	1538
26305 E L21F	743
26767 M 182T	5092
27205 ORF6 del 3	3459
28278 N del 3	409
28308 N A12G	591
28699 N silent	481
28887 N T205I	925
29397 N K375R	1116
29543 intergenic	1078
	7
	27-
	VSP2027-2
	/SF



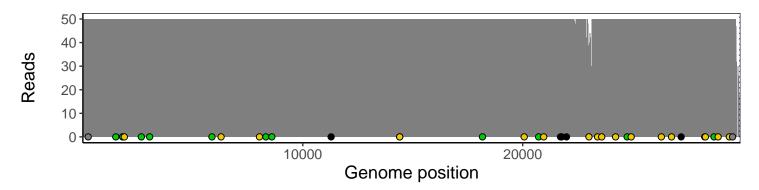
Analyses of individual experiments and composite results

VSP2027-2 | 2021-03-26 | Saline | UPHS-0813 | genomes | single experiment

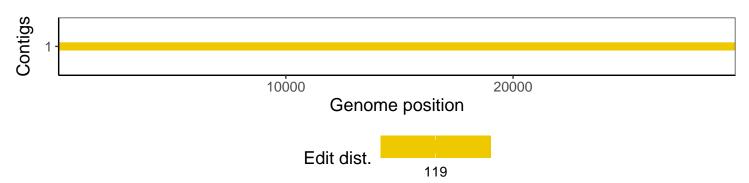
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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