COVID-19 subject UPHS-1353

2021-06-23

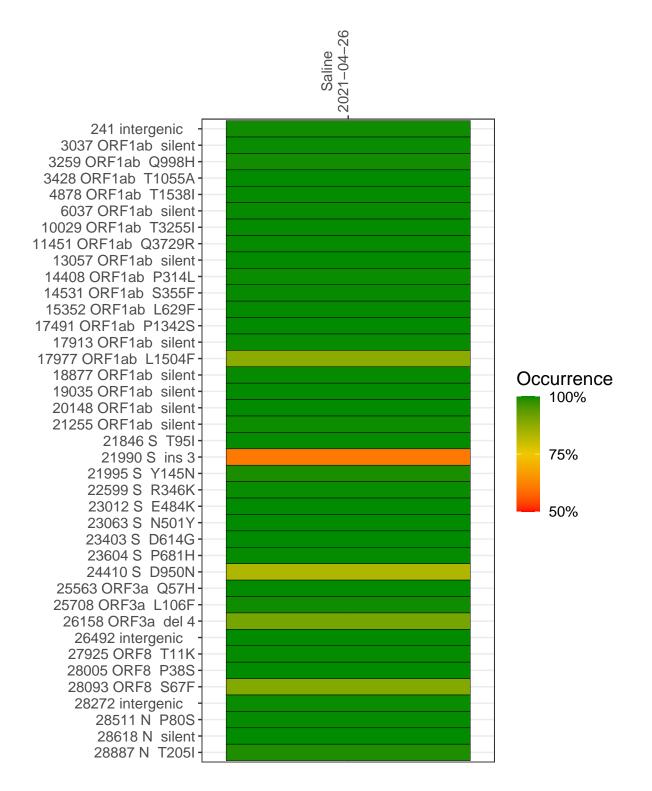
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP2608-1	single experiment	NA	Saline	2021-04-26	29.84	B.1.621	99.7%	99.7%

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.



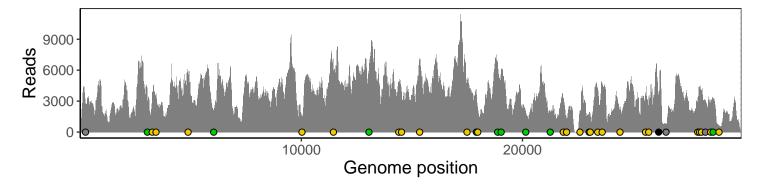
Saline 2021-04-26

	2021-04-20
241 intergenic	2628
3037 ORF1ab silent	3116
3259 ORF1ab Q998H	3533
3428 ORF1ab T1055A	3387
4878 ORF1ab T1538I	5595
6037 ORF1ab silent	2035
10029 ORF1ab T3255I	1520
11451 ORF1ab Q3729R	5437
13057 ORF1ab silent	5842
14408 ORF1ab P314L	4259
14531 ORF1ab S355F	3844
15352 ORF1ab L629F	6849
17491 ORF1ab P1342S	4965
17913 ORF1ab silent	3629
17977 ORF1ab L1504F	2068
18877 ORF1ab silent	6411
19035 ORF1ab silent	4724
20148 ORF1ab silent	2029
21255 ORF1ab silent	1847
21846 S T95I	3025
21990 S ins 3	1191
21995 S Y145N	734
22599 S R346K	1966
23012 S E484K	1477
23063 S N501Y	1669
23403 S D614G	4144
23604 S P681H	4544
24410 S D950N	2174
25563 ORF3a Q57H	3632
25708 ORF3a L106F	2466
26158 ORF3a del 4	2931
26492 intergenic	684
27925 ORF8 T11K	3079
28005 ORF8 P38S	4208
28093 ORF8 S67F	3732
28272 intergenic	3007
28511 N P80S	2737
28618 N silent	2496
28887 N T205I	612
	Ī
	∞

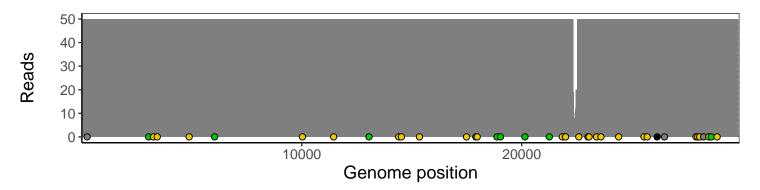
Analyses of individual experiments and composite results

VSP2608-1 | 2021-04-26 | Saline | UPHS-1353 | genomes | single experiment

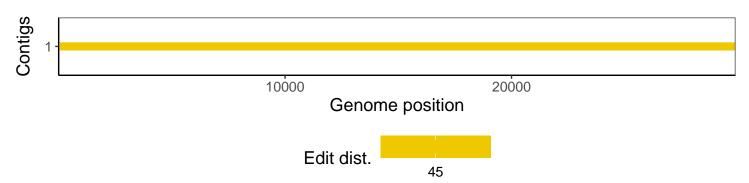
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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