COVID-19 subject UPHS-0596

2021-06-23

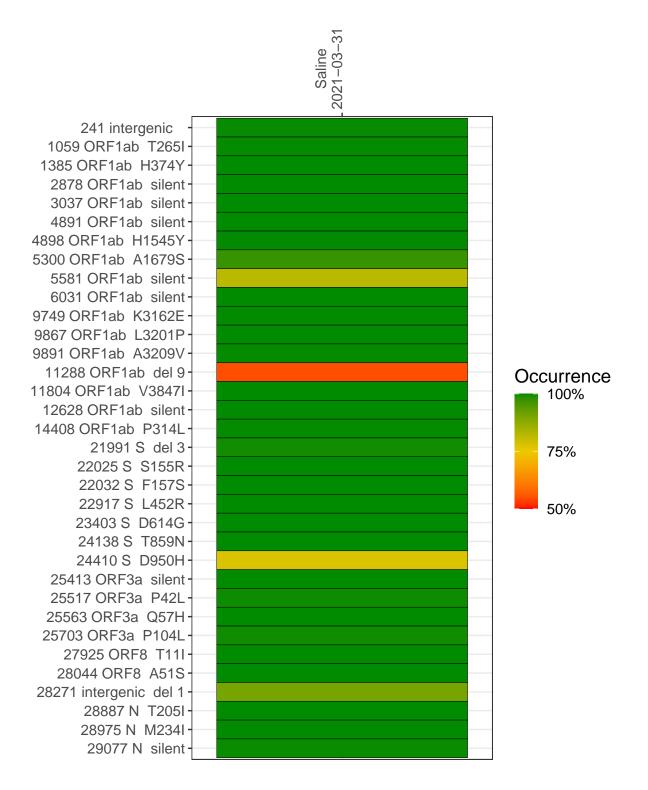
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1781-1	single experiment	NA	Saline	2021-03-31	21.69	B.1.526	98.9%	98.9%

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 2021-03-31

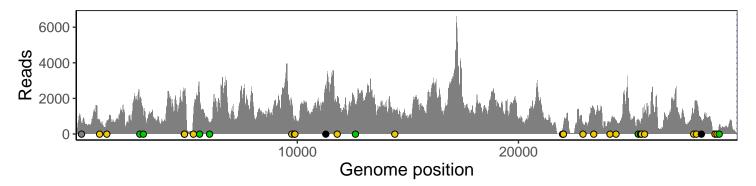
241 intergenic	578
1059 ORF1ab T265I	757
1385 ORF1ab H374Y	711
2878 ORF1ab silent	1659
3037 ORF1ab silent	1011
4891 ORF1ab silent	2183
4898 ORF1ab H1545Y	1996
5300 ORF1ab A1679S	975
5581 ORF1ab silent	2207
6031 ORF1ab silent	582
9749 ORF1ab K3162E	1831
9867 ORF1ab L3201P	702
9891 ORF1ab A3209V	981
11288 ORF1ab del 9	1336
11804 ORF1ab V3847I	1716
12628 ORF1ab silent	1584
14408 ORF1ab P314L	1149
21991 S del 3	214
22025 S S155R	413
22032 S F157S	471
22917 S L452R	548
23403 S D614G	1452
24138 S T859N	641
24410 S D950H	713
25413 ORF3a silent	813
25517 ORF3a P42L	555
25563 ORF3a Q57H	842
25703 ORF3a P104L	1092
27925 ORF8 T11I	1192
28044 ORF8 A51S	1708
28271 intergenic del 1	766
28887 N T205I	173
28975 N M234I	144
29077 N silent	480
	<u></u>
	VSP1781-1
	7
	*



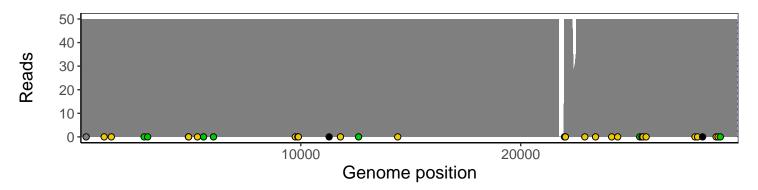
Analyses of individual experiments and composite results

VSP1781-1 | 2021-03-31 | Saline | UPHS-0596 | genomes | single experiment

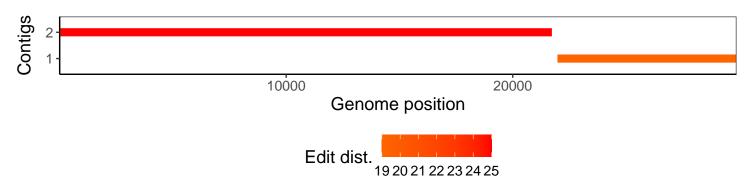
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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