COVID-19 subject UPHS-0048

2021-06-23

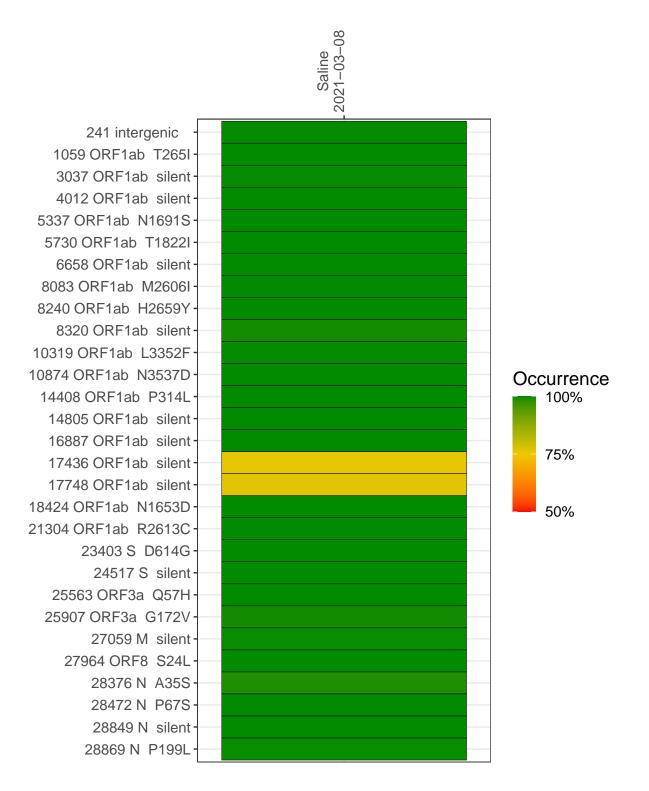
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP0980-1	single experiment	NA	Saline	2021-03-08	29.95	B.1.2	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 2021-03-08

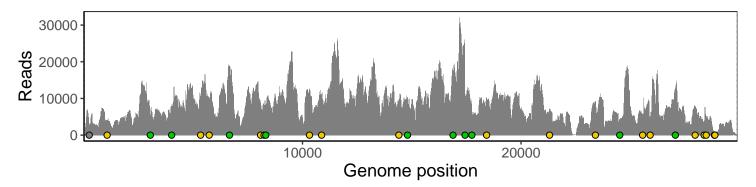
	2021-03-00
241 intergenic	2934
1059 ORF1ab T265I	3988
3037 ORF1ab silent	5777
4012 ORF1ab silent	78 15
5337 ORF1ab N1691S	10629
5730 ORF1ab T1822I	10087
6658 ORF1ab silent	18861
8083 ORF1ab M2606I	8596
8240 ORF1ab H2659Y	8125
8320 ORF1ab silent	7611
10319 ORF1ab L3352F	11441
10874 ORF1ab N3537D	11232
14408 ORF1ab P314L	9682
14805 ORF1ab silent	10175
16887 ORF1ab silent	17529
17436 ORF1ab silent	11232
17748 ORF1ab silent	6160
18424 ORF1ab N1653D	9739
21304 ORF1ab R2613C	7039
23403 S D614G	9154
24517 S silent	4234
25563 ORF3a Q57H	5814
25907 ORF3a G172V	4018
27059 M silent	11488
27964 ORF8 S24L	7975
28376 N A35S	3200
28472 N P67S	5027
28849 N silent	992
28869 N P199L	978
	80-1
	∞



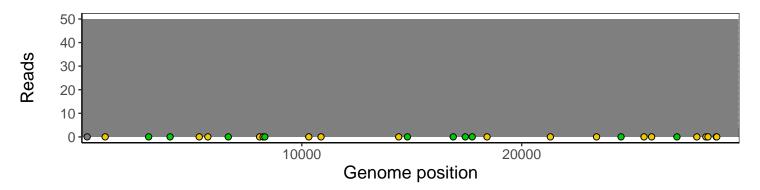
Analyses of individual experiments and composite results

VSP0980-1 | 2021-03-08 | Saline | UPHS-0048 | genomes | single experiment

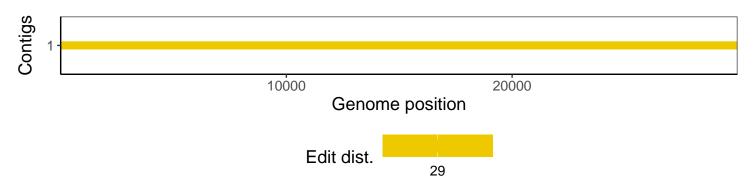
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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				