COVID-19 subject UPHS-0067

2021-03-25

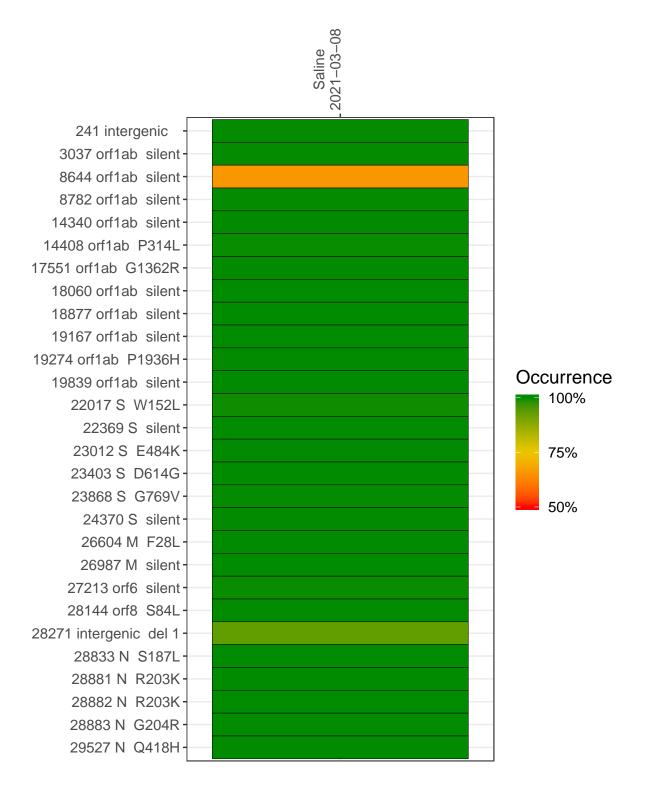
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP0999-1	single experiment	NA	Saline	2021-03-08	29.92	R.1	99.9%	99.8%

Variants shared across samples

The heat map below shows how variants (reference genome USA-WA1-2020) 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

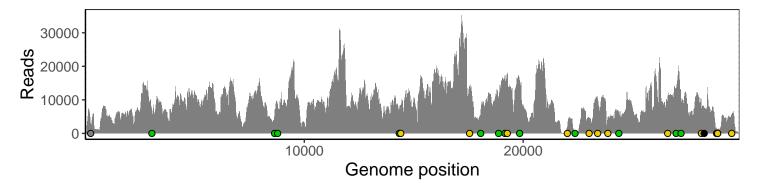
	2021-03-08				
241 intergenic	2813				
3037 orf1ab silent	5969				
8644 orf1ab silent	7051				
8782 orf1ab silent	8341				
14340 orf1ab silent	7064				
14408 orf1ab P314L	8401				
17551 orf1ab G1362R	13584				
18060 orf1ab silent	6474				
18877 orf1ab silent	10241				
19167 orf1ab silent	17004				
19274 orf1ab P1936H	16275				
19839 orf1ab silent	15373				
22017 S W152L	1976				
22369 S silent	105				
23012 S E484K	4270				
23403 S D614G	10660				
23868 S G769V	4840				
24370 S silent	6525				
26604 M F28L	10539				
26987 M silent	11887				
27213 orf6 silent	8121				
28144 orf8 S84L	7732				
28271 intergenic del 1	7406				
28833 N S187L	1445				
28881 N R203K	991				
28882 N R203K	987				
28883 N G204R	992				
29527 N Q418H	5165				
	VSP0999-1				



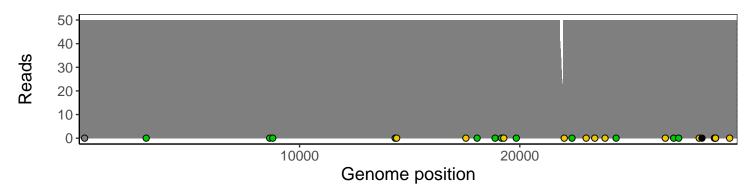
Analyses of individual experiments and composite results

$VSP0999-1 \mid 2021-03-08 \mid Saline \mid UPHS-0067 \mid genomes \mid single \; experiment$

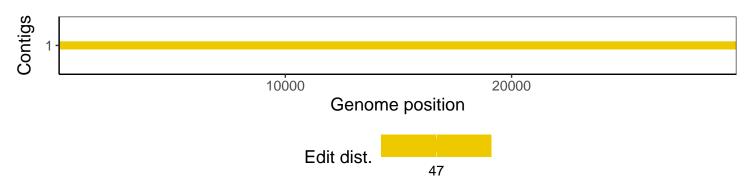
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.3.3
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	3.0.0
tidyverse	1.2.1
ShortRead	1.34.2
$\operatorname{GenomicAlignments}$	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
$\operatorname{GenomeInfoDb}$	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