COVID-19 subject UPHS-0036

2021-03-25

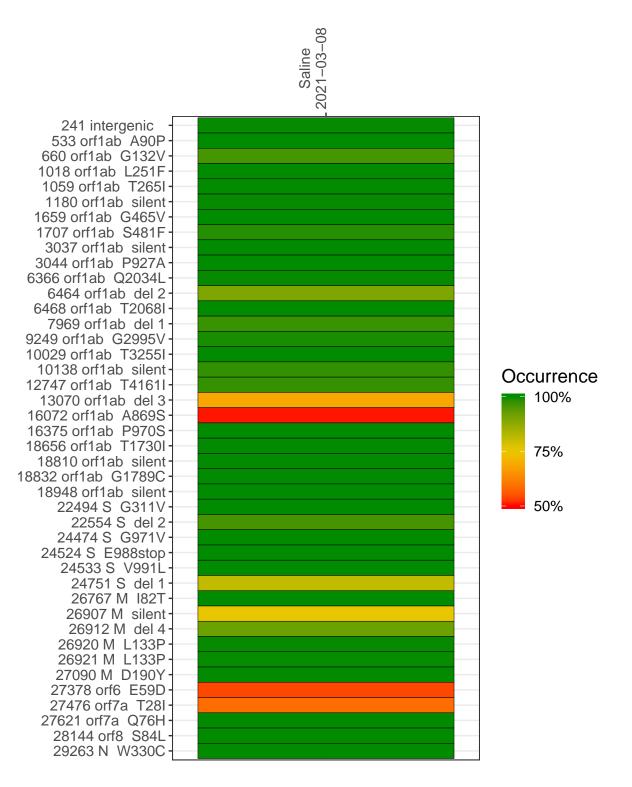
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP0968-1	single experiment	NA	Saline	2021-03-08	1.67	NA	38.6%	37.7%

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	2045
533 orf1ab A90P	735
660 orf1ab G132V	255
1018 orf1ab L251F	2459
1059 orf1ab T265I	2777
1180 orf1ab silent	2744
1659 orf1ab G465V	5572
1707 orf1ab S481F	7827
3037 orf1ab silent	3826
3044 orf1ab P927A	3592
6366 orf1ab Q2034L	9678
6464 orf1ab del 2	6117
6468 orf1ab T2068I	5595
7969 orf1ab del 1	3446
9249 orf1ab G2995V	3466
10029 orf1ab T3255I	495
10138 orf1ab silent	132
12747 orf1ab T4161I	3576
13070 orf1ab del 3	8789
16072 orf1ab A869S	31650
16375 orf1ab P970S	5023
18656 orf1ab T1730l	2282
18810 orf1ab silent	5238
18832 orf1ab G1789C	5588
18948 orf1ab silent	1062
22494 S G311V	32
22554 S del 2	26
24474 S G971V	4535
24524 S E988stop	6334
24533 S V991L	6314
24751 S del 1	2198
26767 M 182T	16176
26907 M silent	3272
26912 M del 4	2429
26920 M L133P	2242
26921 M L133P	2329
27090 M D190Y	7121
27378 orf6 E59D	15117
27476 orf7a T28I	10733
27621 orf7a Q76H	2341
28144 orf8 S84L	2802
29263 N W330C	5322
	-

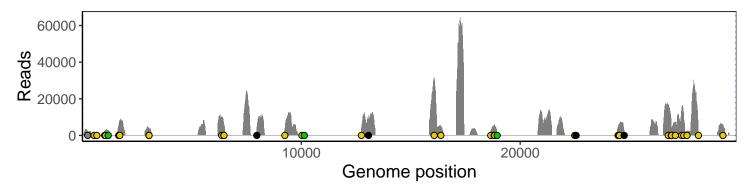


SP0968-1

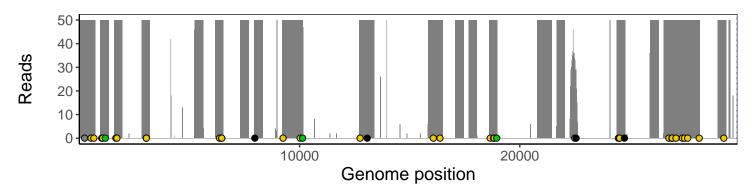
Analyses of individual experiments and composite results

$VSP0968-1 \mid 2021-03-08 \mid Saline \mid UPHS-0036 \mid genomes \mid single \; experiment$

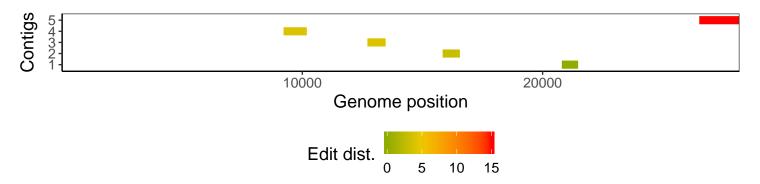
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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
SummarizedExperiment	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