COVID-19 subject UPHS-0316

2021-04-17

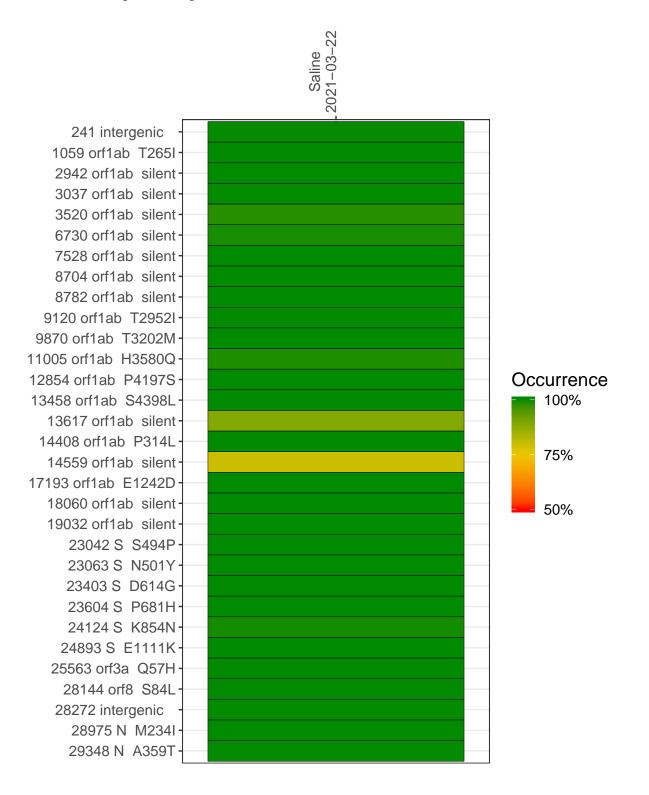
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1361-1	single experiment	NA	Saline	2021-03-22	29.61	B.1	99.1%	99.1%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/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 2021-03-22

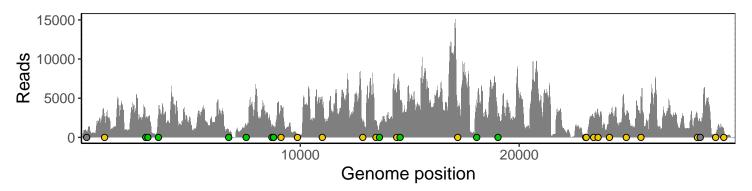
	2021-03-22
241 intergenic	727
1059 orf1ab T265I	2015
2942 orf1ab silent	2055
3037 orf1ab silent	2203
3520 orf1ab silent	2821
6730 orf1ab silent	887
7528 orf1ab silent	1548
8704 orf1ab silent	1604
8782 orf1ab silent	1344
9120 orf1ab T2952I	4114
9870 orf1ab T3202M	143
11005 orf1ab H3580Q	4646
12854 orf1ab P4197S	1765
13458 orf1ab S4398L	1854
13617 orf1ab silent	1415
14408 orf1ab P314L	5256
14559 orf1ab silent	4709
17193 orf1ab E1242D	4550
18060 orf1ab silent	3708
19032 orf1ab silent	2634
23042 S S494P	375
23063 S N501Y	371
23403 S D614G	2541
23604 S P681H	3177
24124 S K854N	2804
24893 S E1111K	2747
25563 orf3a Q57H	5001
28144 orf8 S84L	2536
28272 intergenic	2046
28975 N M234I	422
29348 N A359T	901
	7
	361
	VSP1361-1
	× ×



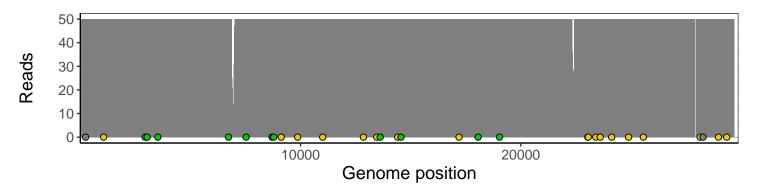
Analyses of individual experiments and composite results

VSP1361-1 | 2021-03-22 | Saline | UPHS-0316 | genomes | single experiment

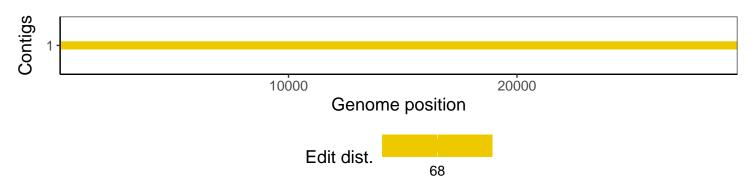
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.3.8
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
${\it Genomic Alignments}$	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