# COVID-19 subject UPHS-0283

2021-04-17

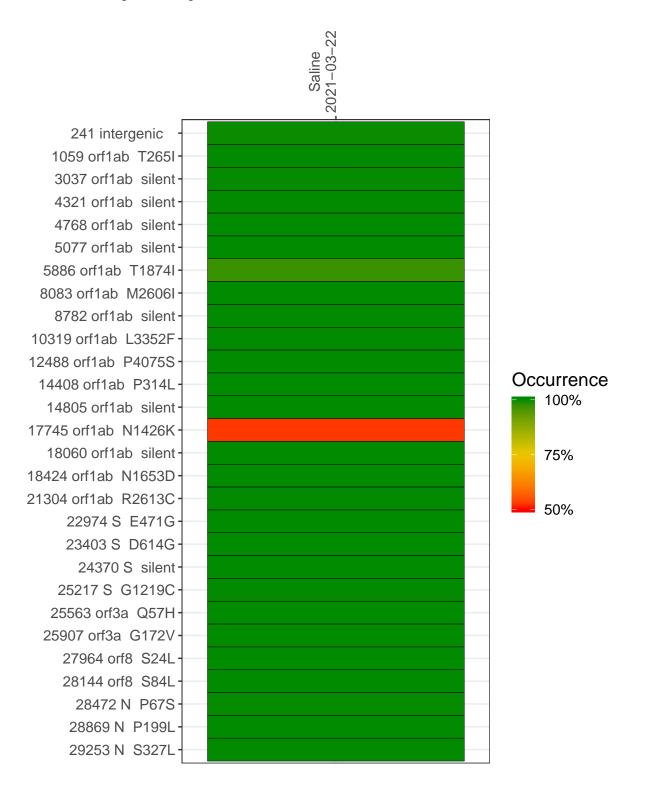
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with > 90% sequence coverage.

Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1328-1	single experiment	NA	Saline	2021-03-22	29.70	B.1.2	99.3%	99.2%

#### 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

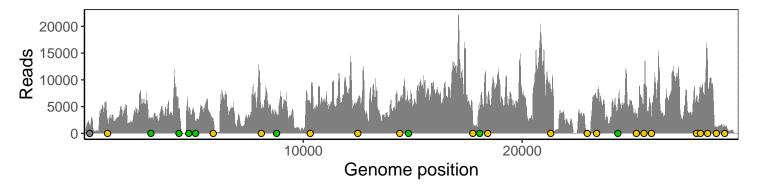
	2021-03-22
241 intergenic	1504
1059 orf1ab T265I	1945
3037 orf1ab silent	2652
4321 orf1ab silent	3893
4768 orf1ab silent	4782
5077 orf1ab silent	2303
5886 orf1ab T1874I	2800
8083 orf1ab M2606I	4146
8782 orf1ab silent	3470
10319 orf1ab L3352F	5587
12488 orf1ab P4075S	5246
14408 orf1ab P314L	5429
14805 orf1ab silent	5918
17745 orf1ab N1426K	3511
18060 orf1ab silent	4371
18424 orf1ab N1653D	6091
21304 orf1ab R2613C	7976
22974 S E471G	402
23403 S D614G	6470
24370 S silent	5557
25217 S G1219C	3488
25563 orf3a Q57H	8799
25907 orf3a G172V	3059
27964 orf8 S24L	6750
28144 orf8 S84L	8572
28472 N P67S	13334
28869 N P199L	1116
29253 N S327L	1259
	<del></del>



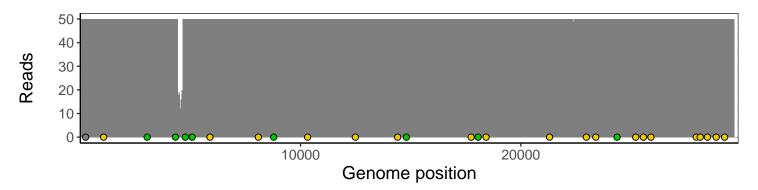
### Analyses of individual experiments and composite results

#### $VSP1328-1 \mid 2021-03-22 \mid Saline \mid UPHS-0283 \mid genomes \mid single \; experiment$

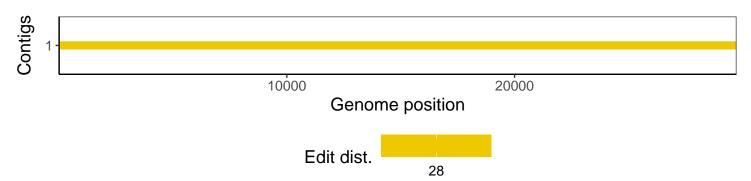
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.3.8
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