# COVID-19 subject UPHS-0265

2021-05-05

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)
VSP1310-1	single experiment	NA	Saline	2021-03-08	21.46	B.1.2	98.6%	98.5%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/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.
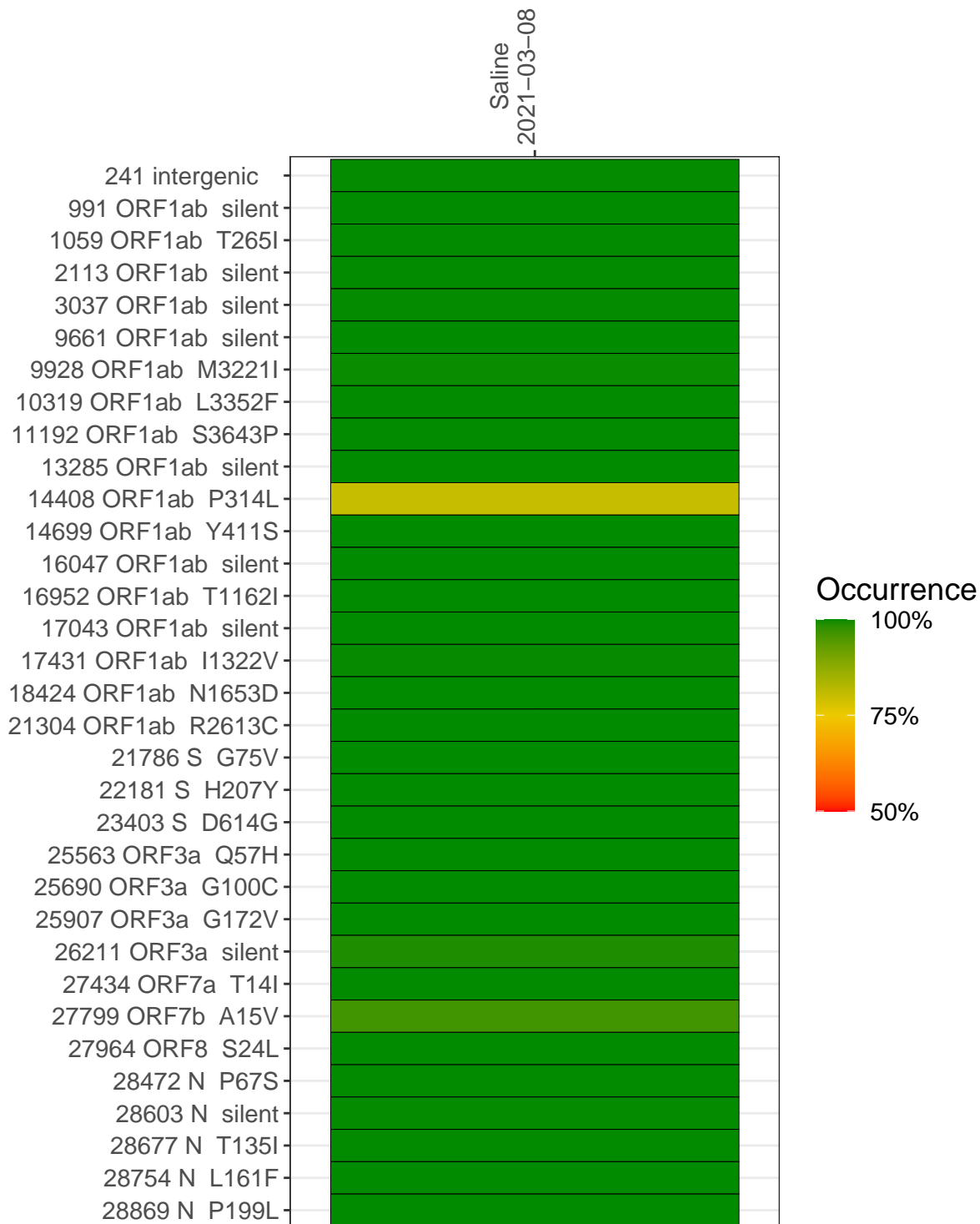
Saline
2021-03-08

| Position	VSP1310-1
241 intergenic	1113
991 ORF1ab silent	4751
1059 ORF1ab T265I	1397
2113 ORF1ab silent	1379
3037 ORF1ab silent	858
9661 ORF1ab silent	353
9928 ORF1ab M3221I	424
10319 ORF1ab L3352F	3165
11192 ORF1ab S3643P	5085
13285 ORF1ab silent	6080
14408 ORF1ab P314L	2160
14699 ORF1ab Y411S	572
16047 ORF1ab silent	2316
16952 ORF1ab T1162I	7064
17043 ORF1ab silent	6078
17431 ORF1ab I1322V	8468
18424 ORF1ab N1653D	837
21304 ORF1ab R2613C	6204
21786 S G75V	572
22181 S H207Y	235
23403 S D614G	6922
25563 ORF3a Q57H	5221
25690 ORF3a G100C	2877
25907 ORF3a G172V	2414
26211 ORF3a silent	11173
27434 ORF7a T14I	2533
27799 ORF7b A15V	1157
27964 ORF8 S24L	354
28472 N P67S	7240
28603 N silent	5361
28677 N T135I	4585
28754 N L161F	3007
28869 N P199L	975

Base change

- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

Analyses of individual experiments and composite results

## VSP1310-1 | 2021-03-08 | Saline | UPHS-0265 | 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.
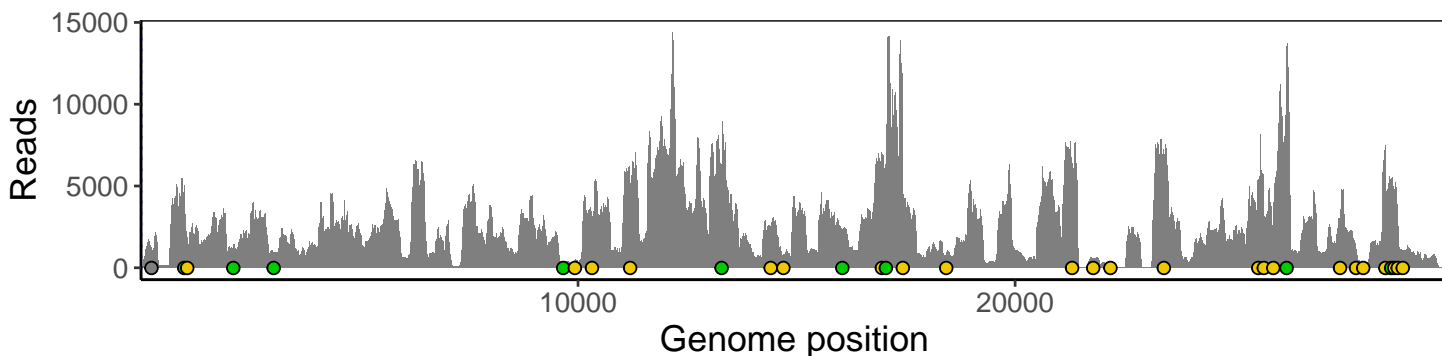
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
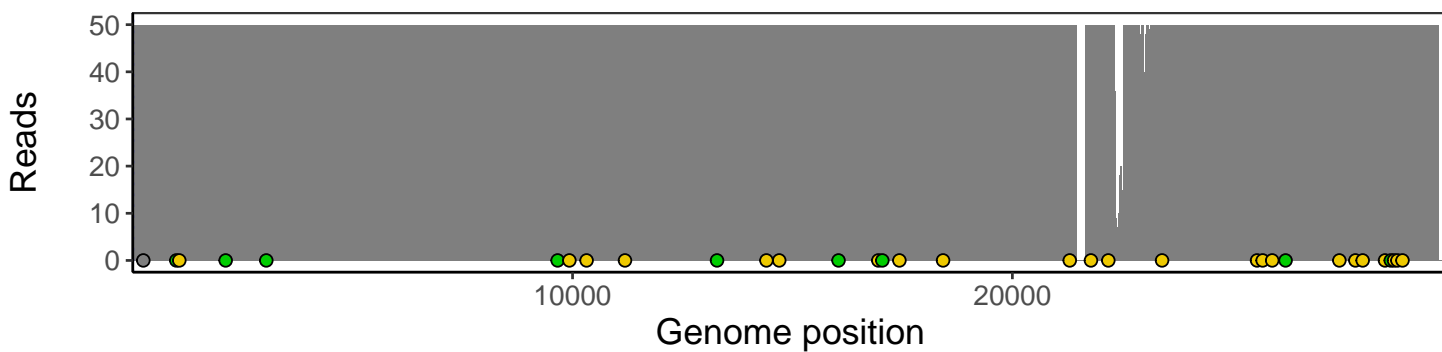
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
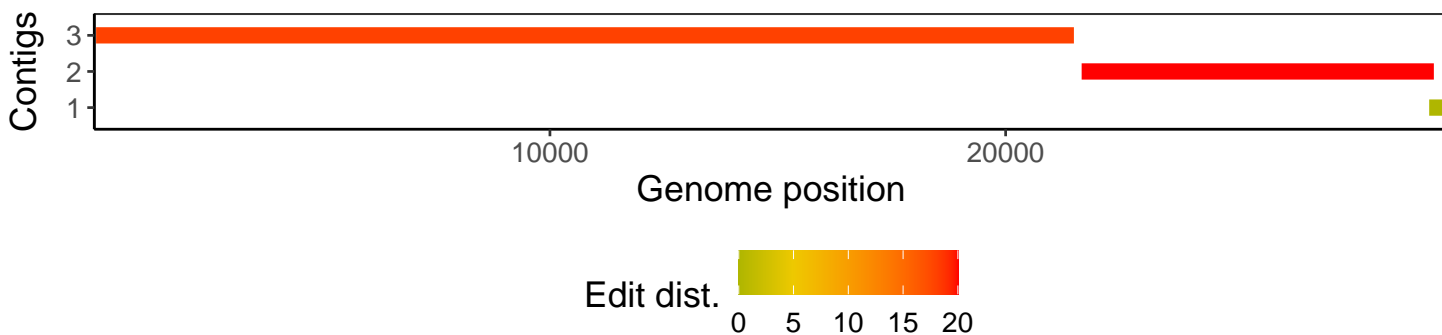
Software environment

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
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
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