COVID-19 subject UPHS-0302

2021-06-23

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)
VSP1347-1	single experiment	NA	Saline	2021-03-14	29.79	B.1.525	99.7%	99.6%

Variants shared across samples

The heat map below shows how variants (reference genome /home/common/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
2021−03−14
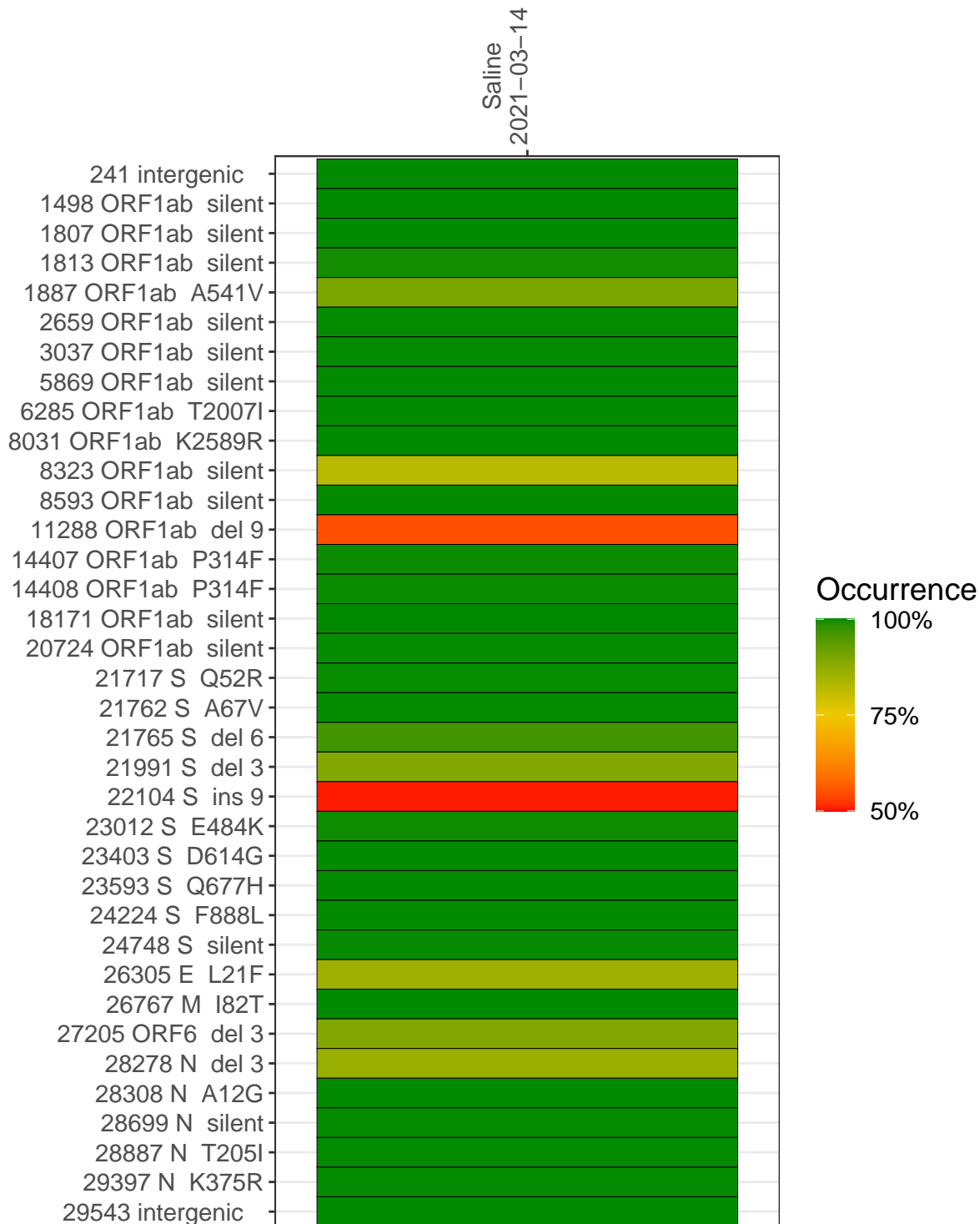
| Position	Value	Base change
241 intergenic	2512	T
1498 ORF1ab silent	4103	T
1807 ORF1ab silent	4842	G
1813 ORF1ab silent	4886	C
1887 ORF1ab A541V	5849	T
2659 ORF1ab silent	9500	A
3037 ORF1ab silent	3820	T
5869 ORF1ab silent	3916	T
6285 ORF1ab T2007I	11668	T
8031 ORF1ab K2589R	9344	G
8323 ORF1ab silent	7886	Expected/C
8593 ORF1ab silent	4469	C
11288 ORF1ab del 9	4730	Ins/Del
14407 ORF1ab P314F	3131	T
14408 ORF1ab P314F	3193	T
18171 ORF1ab silent	4684	T
20724 ORF1ab silent	10058	G
21717 S Q52R	4054	G
21762 S A67V	2033	T
21765 S del 6	1988	Ins/Del
21991 S del 3	1971	Ins/Del
22104 S ins 9	4235	Expected
23012 S E484K	300	A
23403 S D614G	9335	G
23593 S Q677H	8864	C
24224 S F888L	7716	C
24748 S silent	13978	T
26305 E L21F	7116	T
26767 M I82T	7605	C
27205 ORF6 del 3	3891	Ins/Del
28278 N del 3	5659	Ins/Del
28308 N A12G	7793	G
28699 N silent	10573	G
28887 N T205I	2279	T
29397 N K375R	4095	G
29543 intergenic	3713	T

VSP1347−1

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

3

Analyses of individual experiments and composite results

**VSP1347-1 | 2021-03-14 | Saline | UPHS-0302 | 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.
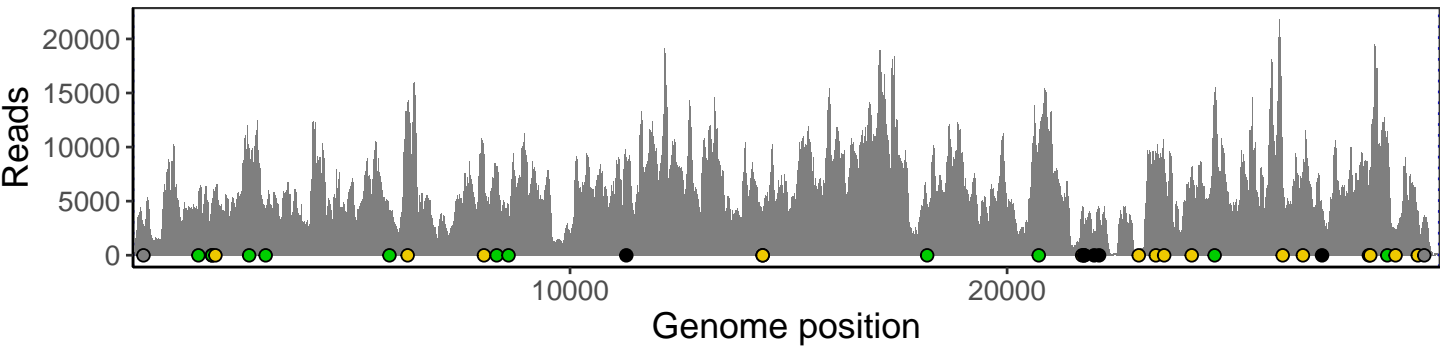
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
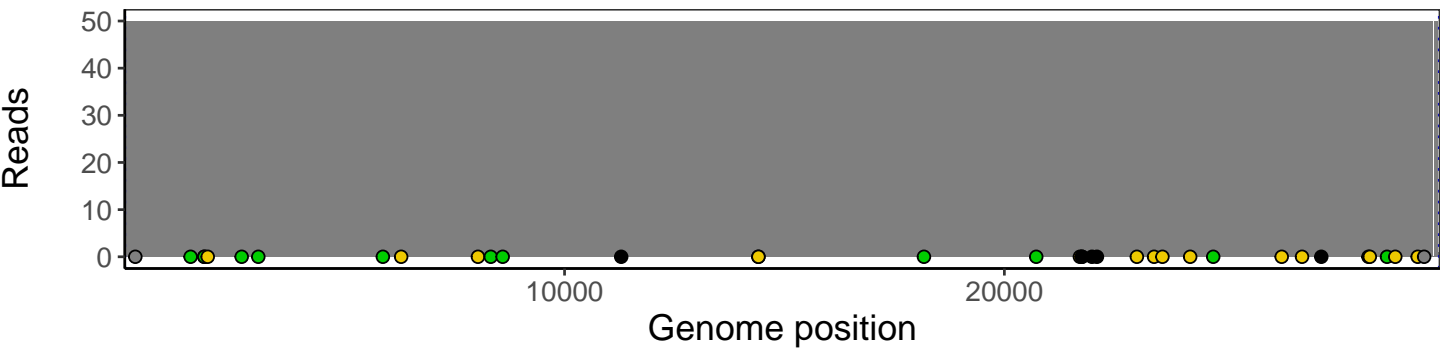
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
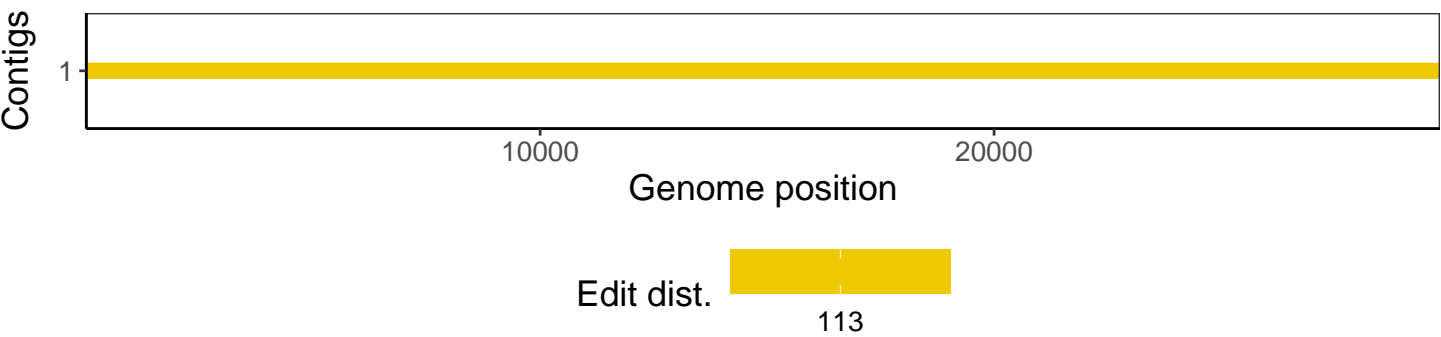
Software environment

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	3.1.3
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	3.3.3
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