COVID-19 subject UPHS-0026

2021-03-25

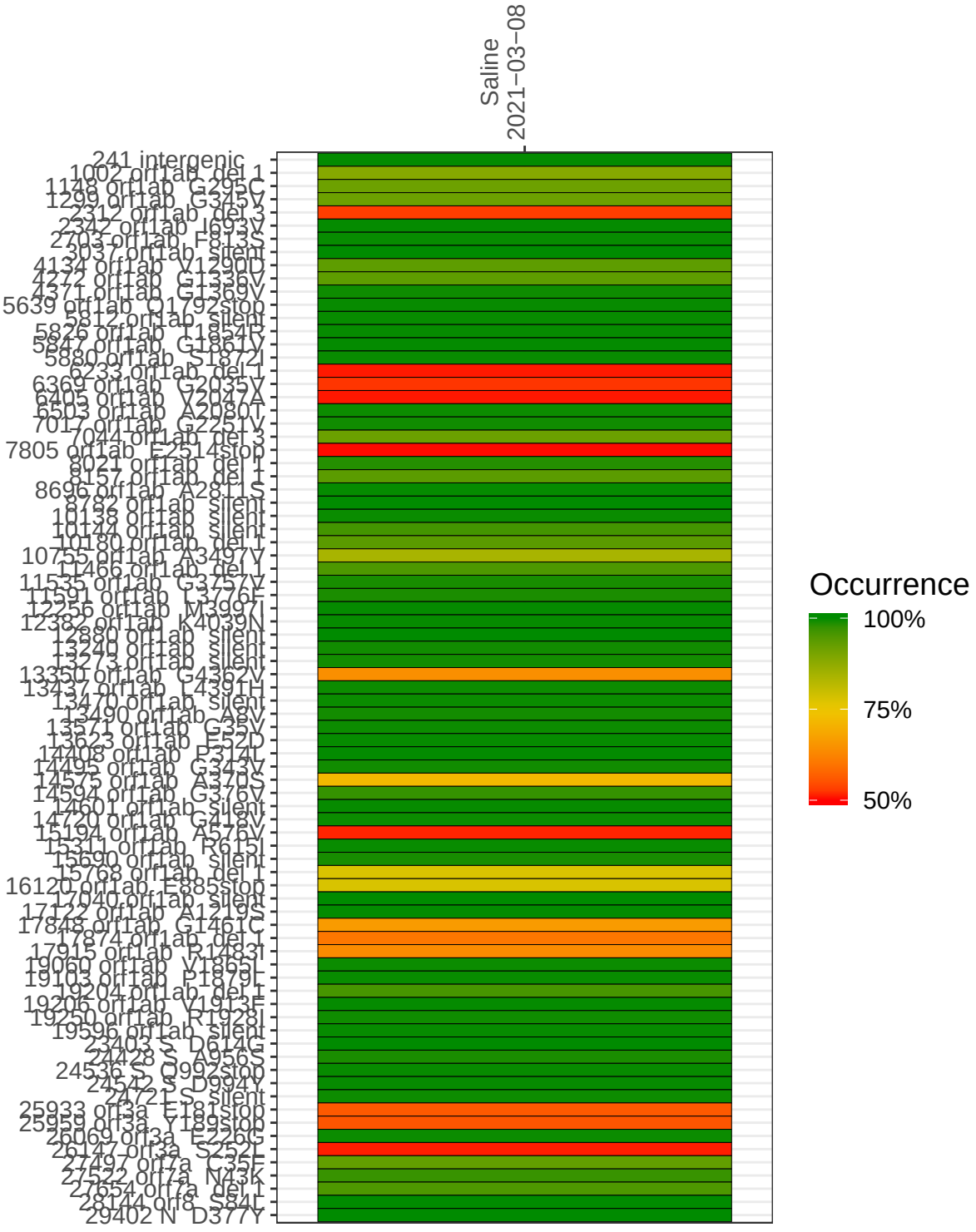
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

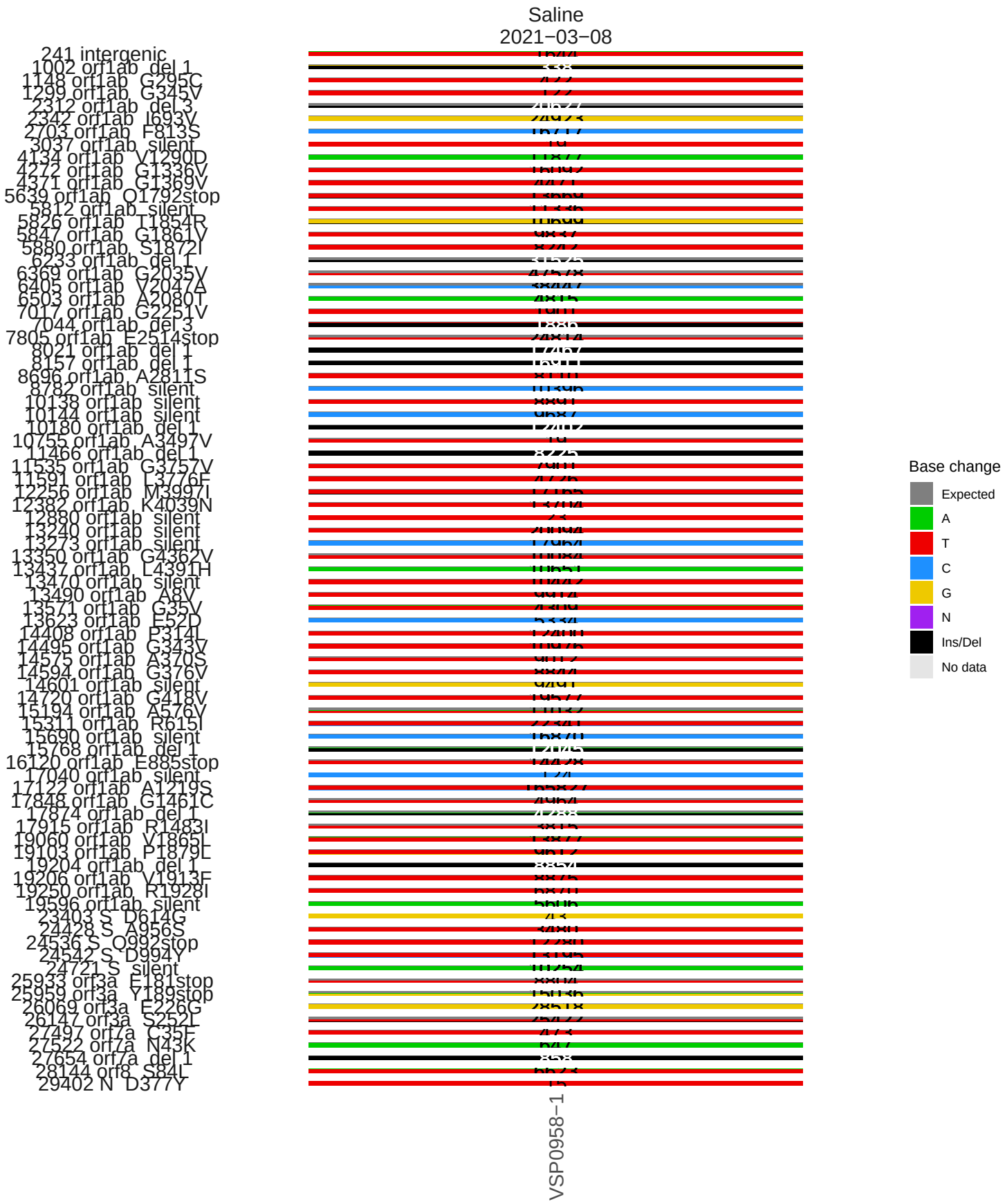
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0958-1	single experiment	NA	Saline	2021-03-08	8.44	B.1.1.304	97.9%	92.0%

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.

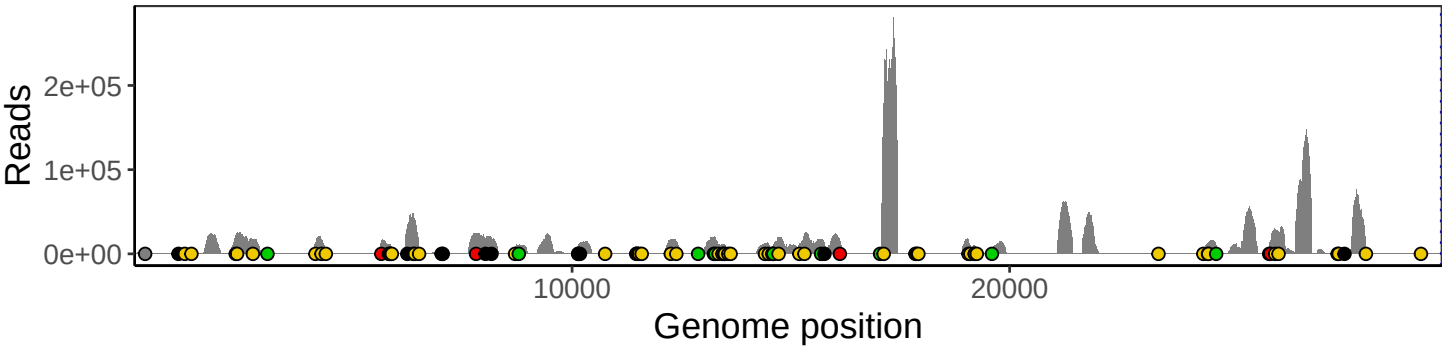



Saline
2021−03−08
241 intergenic
1002 orf1ab del 1
1148 orf1ab G295C
1299 orf1ab G345V
2312 orf1ab del 3
2342 orf1ab I693V
2703 orf1ab F813S
3037 orf1ab silent
4134 orf1ab V1290D
4272 orf1ab G1336V
4371 orf1ab G1369V
5639 orf1ab Q1792stop
5812 orf1ab silent
5826 orf1ab T1854R
5847 orf1ab G1861V
5880 orf1ab S1872I
6233 orf1ab del 1
6369 orf1ab G2035V
6405 orf1ab V2047A
6503 orf1ab A2080T
7017 orf1ab G2251V
7044 orf1ab del 3
7805 orf1ab E2514stop
8021 orf1ab del 1
8157 orf1ab del 1
8696 orf1ab A2811S
8782 orf1ab silent
10138 orf1ab silent
10144 orf1ab silent
10180 orf1ab del 1
10755 orf1ab A3497V
11466 orf1ab del 1
11535 orf1ab G3757V
11591 orf1ab L3776F
12256 orf1ab M3997I
12382 orf1ab K4039N
12880 orf1ab silent
13240 orf1ab silent
13273 orf1ab silent
13350 orf1ab G4362V
13437 orf1ab L4391H
13470 orf1ab silent
13490 orf1ab A8V
13571 orf1ab G35V
13623 orf1ab E52D
14408 orf1ab P314L
14495 orf1ab G343V
14575 orf1ab A370S
14594 orf1ab G376V
14601 orf1ab silent
14720 orf1ab G418V
15194 orf1ab A576V
15311 orf1ab R615I
15690 orf1ab silent
15768 orf1ab del 1
16120 orf1ab E885stop
17040 orf1ab silent
17122 orf1ab A1219S
17848 orf1ab G1461C
17874 orf1ab del 1
17915 orf1ab R1483I
19060 orf1ab V1865L
19103 orf1ab P1879L
19204 orf1ab del 1
19206 orf1ab V1913F
19250 orf1ab R1928I
19596 orf1ab silent
23403 S D614G
24428 S A956S
24536 S Q992stop
24542 S D994Y
24721 S silent
25933 orf3a E181stop
25959 orf3a Y189stop
26069 orf3a E226G
26147 orf3a S252L
27497 orf7a C35F
27522 orf7a N43K
27654 orf7a del 1
28144 orf8 S84L
29402 N D377Y
VSP0958−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

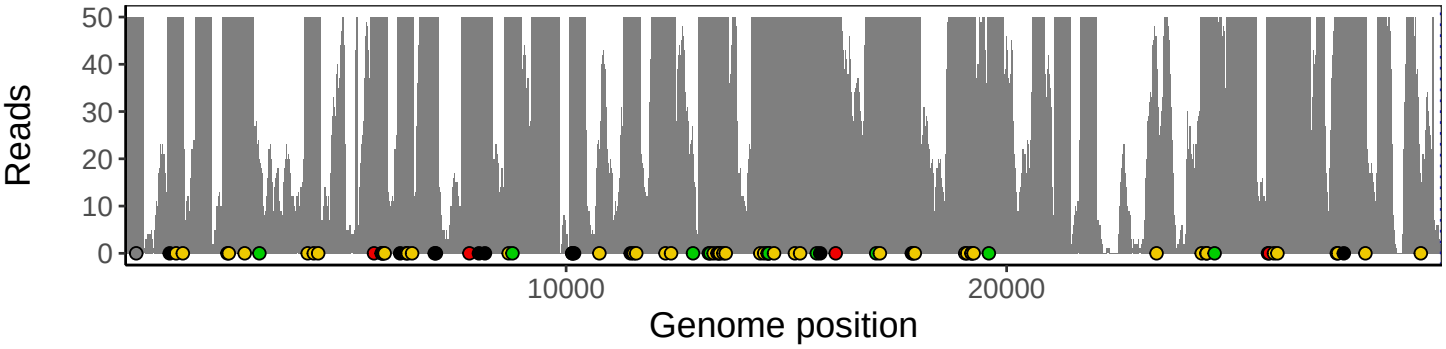
Analyses of individual experiments and composite results

VSP0958-1 | 2021-03-08 | Saline | UPHS-0026 | genomes | single experiment

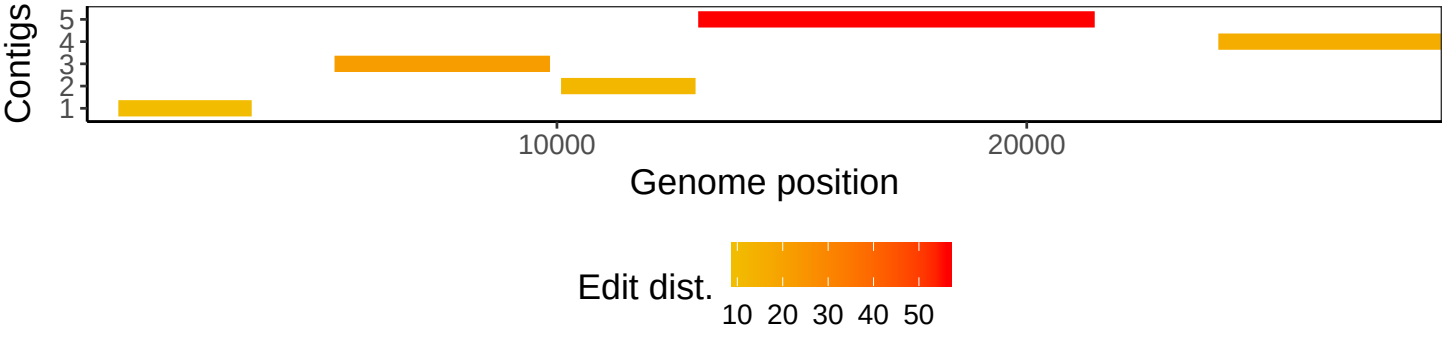
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.3.3
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
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