COVID-19 subject UPHS-0044

2021-03-25

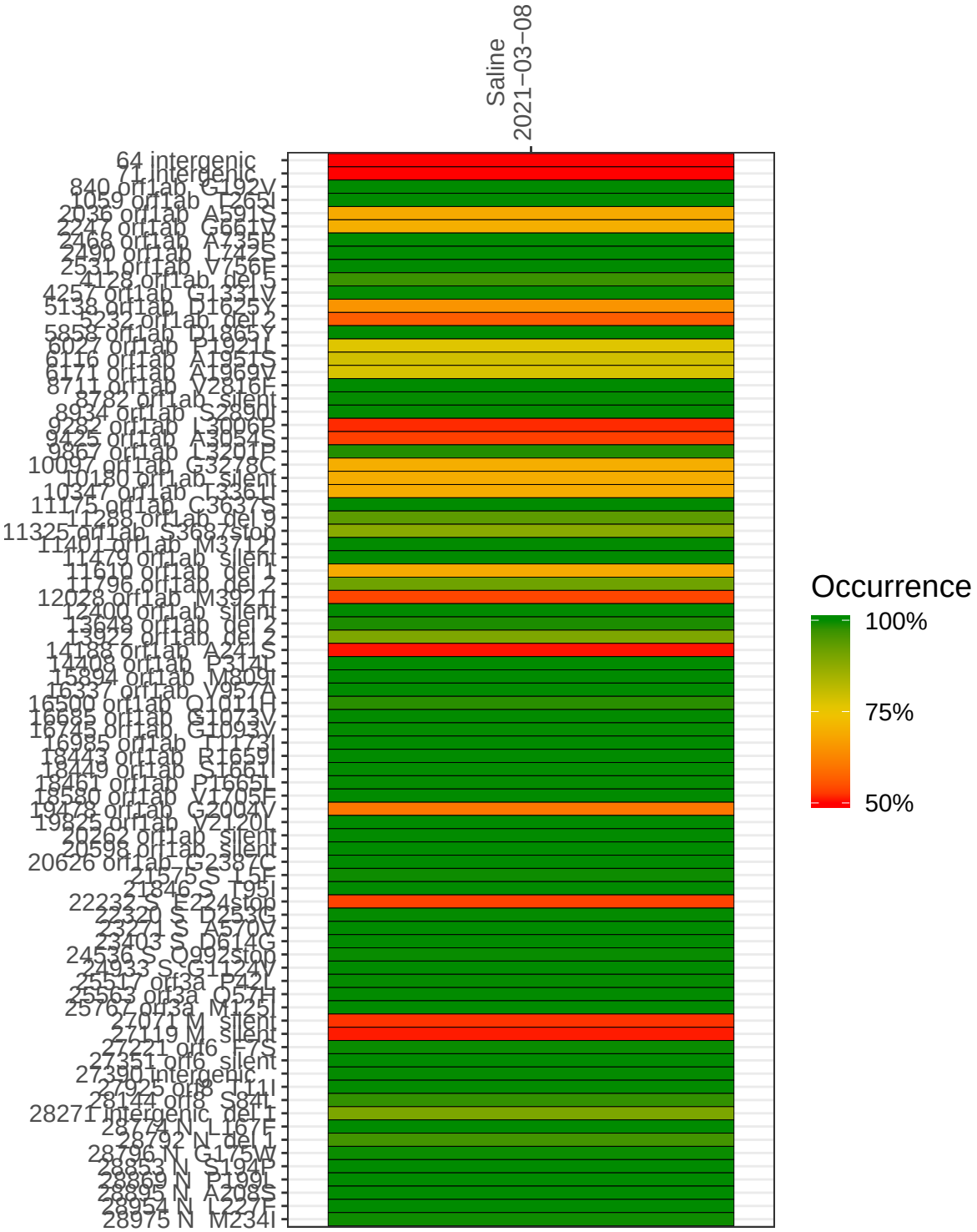
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

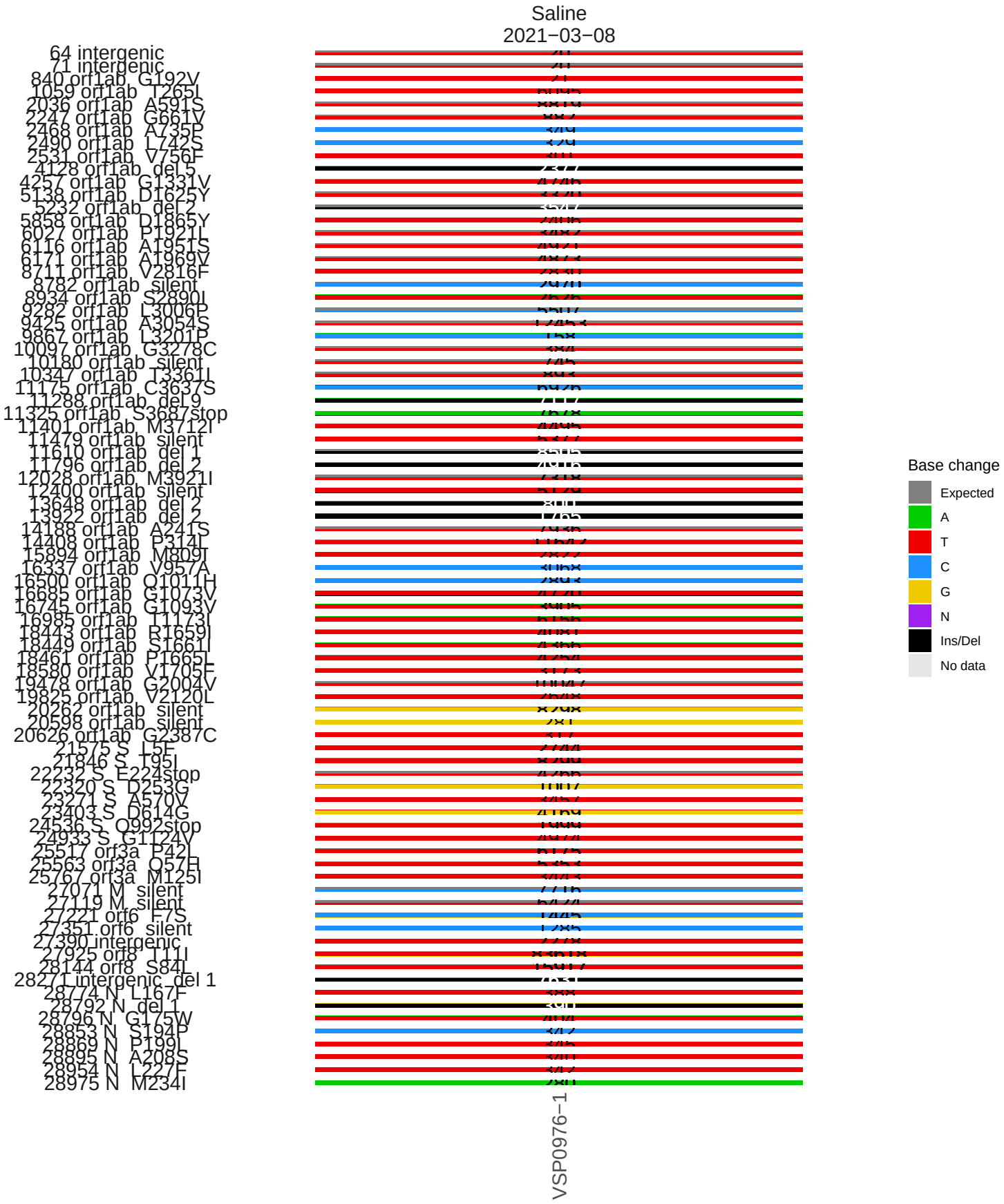
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0976-1	single experiment	NA	Saline	2021-03-08	3.86	NA	74.3%	73.6%

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.

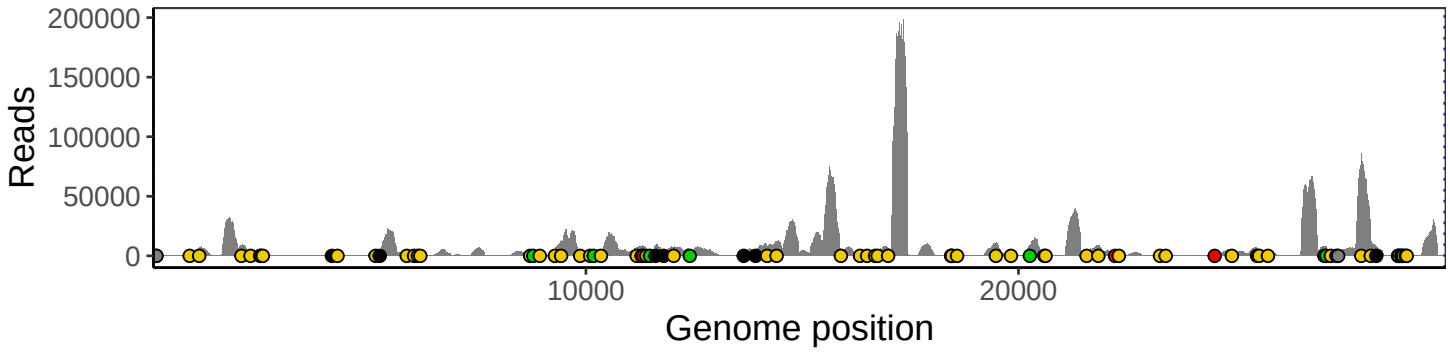



Saline
2021−03−08
64 intergenic
71 intergenic
840 orf1ab G192V
1059 orf1ab T265I
2036 orf1ab A591S
2247 orf1ab G661V
2468 orf1ab A735P
2490 orf1ab L742S
2531 orf1ab V756F
4128 orf1ab del 5
4257 orf1ab G1331V
5138 orf1ab D1625Y
5232 orf1ab del 2
5858 orf1ab D1865Y
6027 orf1ab P1921L
6116 orf1ab A1951S
6171 orf1ab A1969V
8711 orf1ab V2816F
8782 orf1ab silent
8934 orf1ab S2890I
9282 orf1ab L3006P
9425 orf1ab A3054S
9867 orf1ab L3201P
10097 orf1ab G3278C
10180 orf1ab silent
10347 orf1ab T3361I
11175 orf1ab C3637S
11288 orf1ab del 9
11325 orf1ab S3687stop
11401 orf1ab M3712I
11479 orf1ab silent
11610 orf1ab del 1
11796 orf1ab del 2
12028 orf1ab M3921I
12400 orf1ab silent
13648 orf1ab del 2
13922 orf1ab del 2
14188 orf1ab A241S
14408 orf1ab P314L
15894 orf1ab M809I
16337 orf1ab V957A
16500 orf1ab Q1011H
16685 orf1ab G1073V
16745 orf1ab G1093V
16985 orf1ab T1173I
18443 orf1ab R1659I
18449 orf1ab S1661I
18461 orf1ab P1665L
18580 orf1ab V1705F
19478 orf1ab G2004V
19825 orf1ab V2120L
20262 orf1ab silent
20598 orf1ab silent
20626 orf1ab G2387C
21575 S L5F
21846 S T95I
22232 S E224stop
22320 S D253G
23271 S A570V
23403 S D614G
24536 S Q992stop
24933 S G1124V
25517 orf3a P42L
25563 orf3a Q57H
25767 orf3a M125I
27071 M silent
27119 M silent
27221 orf6 F7S
27351 orf6 silent
27390 intergenic
27925 orf8 T11I
28144 orf8 S84L
28271 intergenic del 1
28774 N L167F
28792 N del 1
28796 N G175W
28853 N S194P
28869 N P199L
28895 N A208S
28954 N L227F
28975 N M234I
VSP0976−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

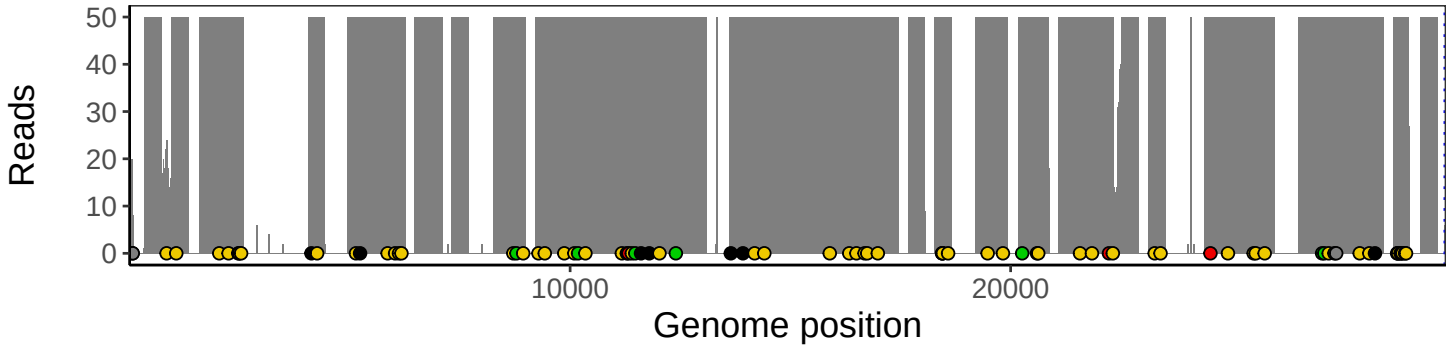
Analyses of individual experiments and composite results

VSP0976-1 | 2021-03-08 | Saline | UPHS-0044 | genomes | single experiment

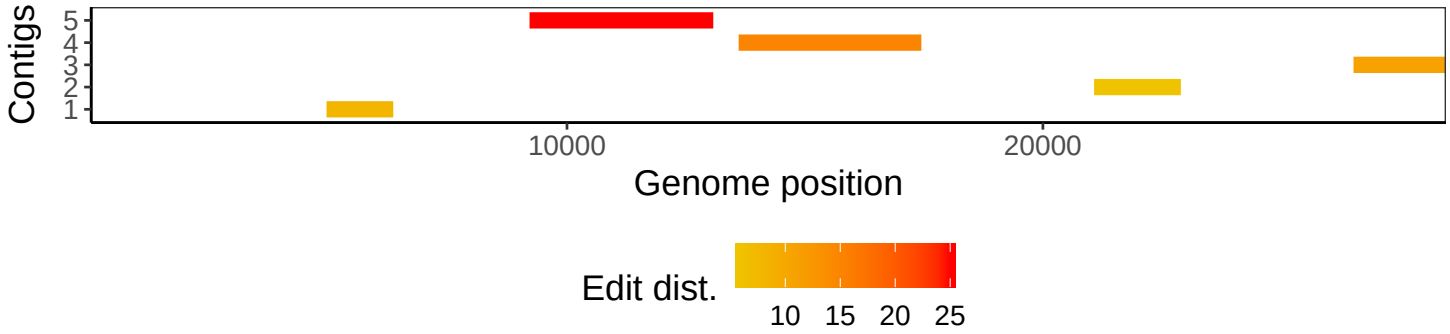
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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