COVID-19 subject UPHS-0088

2021-03-25

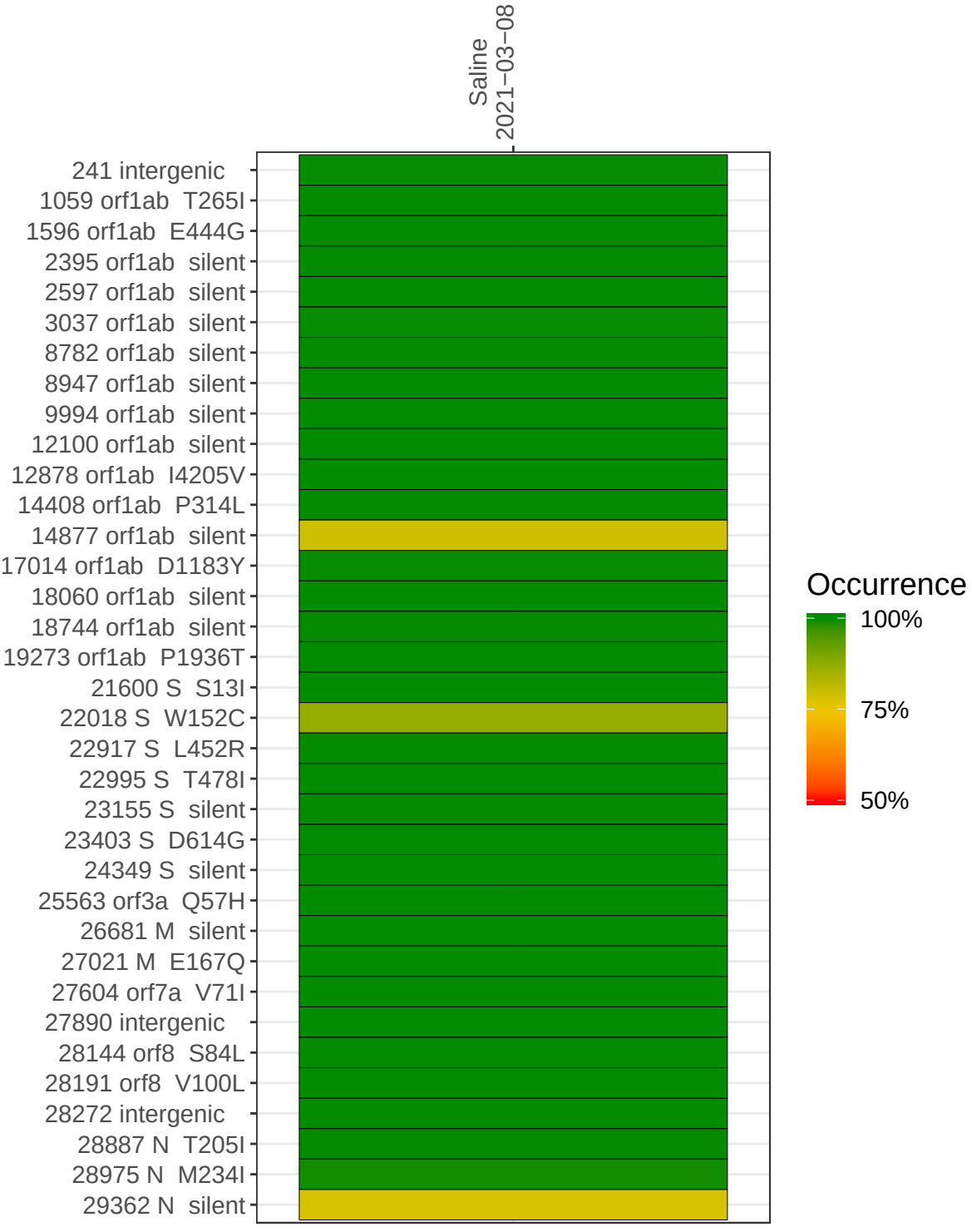
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

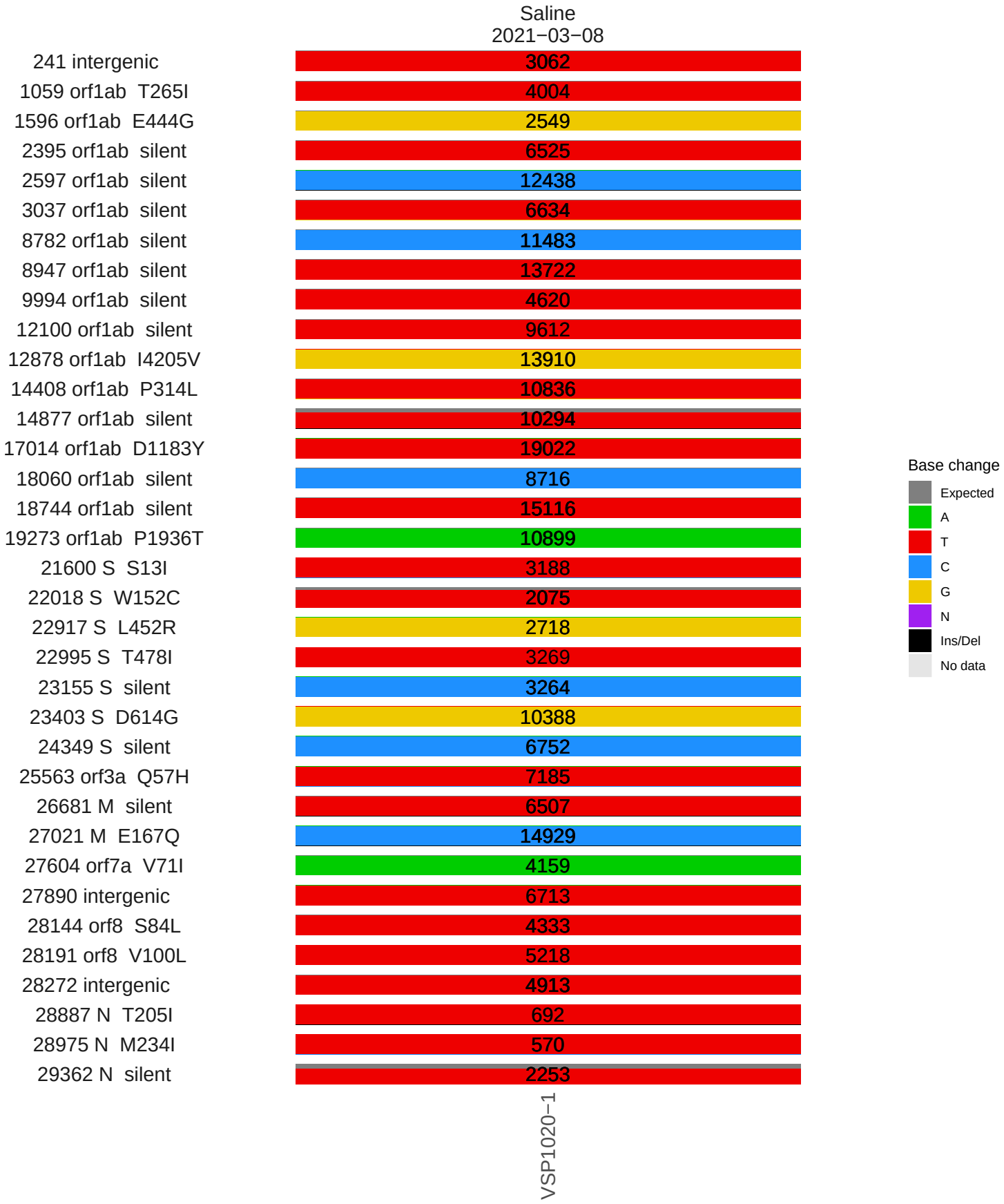
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1020-1	single experiment	NA	Saline	2021-03-08	29.99	B.1.429	99.9%	99.8%

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.

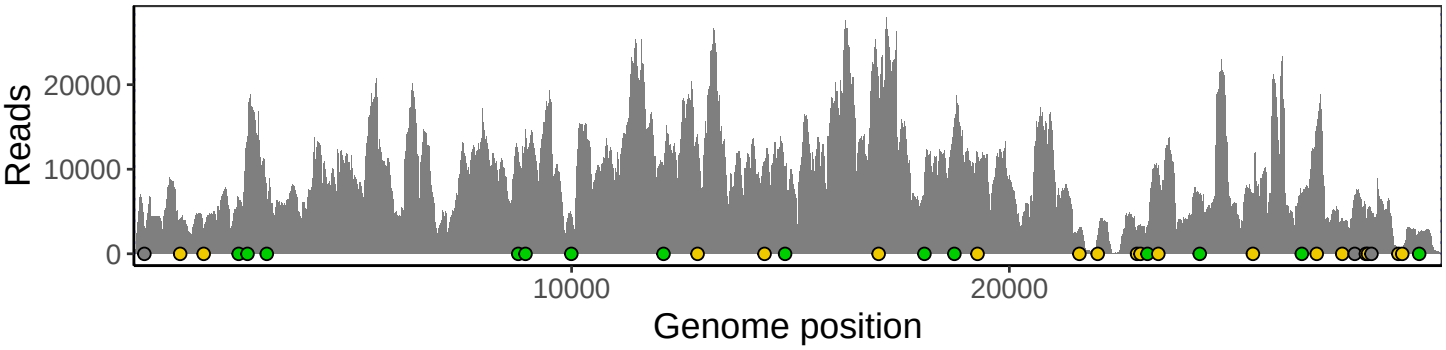



Saline
2021−03−08
241 intergenic
1059 orf1ab T265I
1596 orf1ab E444G
2395 orf1ab silent
2597 orf1ab silent
3037 orf1ab silent
8782 orf1ab silent
8947 orf1ab silent
9994 orf1ab silent
12100 orf1ab silent
12878 orf1ab I4205V
14408 orf1ab P314L
14877 orf1ab silent
17014 orf1ab D1183Y
18060 orf1ab silent
18744 orf1ab silent
19273 orf1ab P1936T
21600 S S13I
22018 S W152C
22917 S L452R
22995 S T478I
23155 S silent
23403 S D614G
24349 S silent
25563 orf3a Q57H
26681 M silent
27021 M E167Q
27604 orf7a V71I
27890 intergenic
28144 orf8 S84L
28191 orf8 V100L
28272 intergenic
28887 N T205I
28975 N M234I
29362 N silent
3062
4004
2549
6525
12438
6634
11483
13722
4620
9612
13910
10836
10294
19022
8716
15116
10899
3188
2075
2718
3269
3264
10388
6752
7185
6507
14929
4159
6713
4333
5218
4913
692
570
2253
VSP1020−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

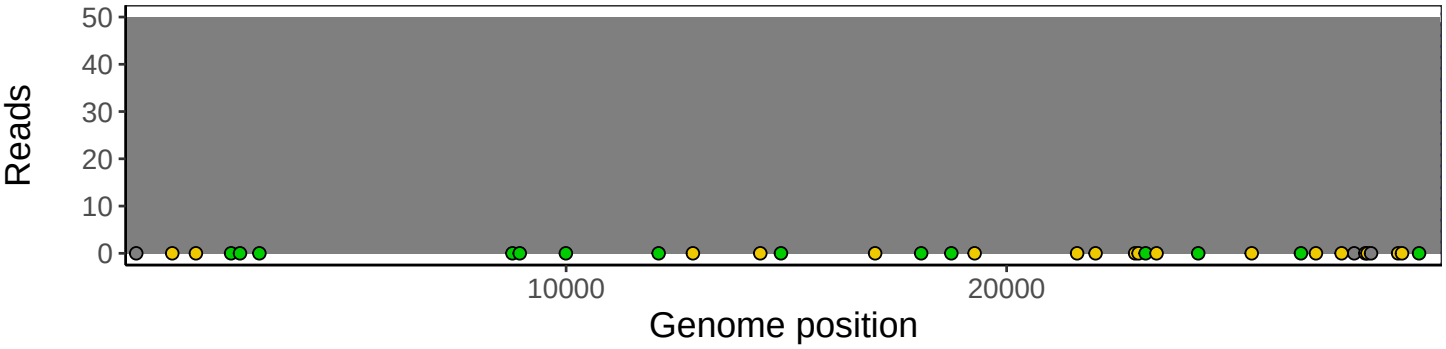
Analyses of individual experiments and composite results

VSP1020-1 | 2021-03-08 | Saline | UPHS-0088 | genomes | single experiment

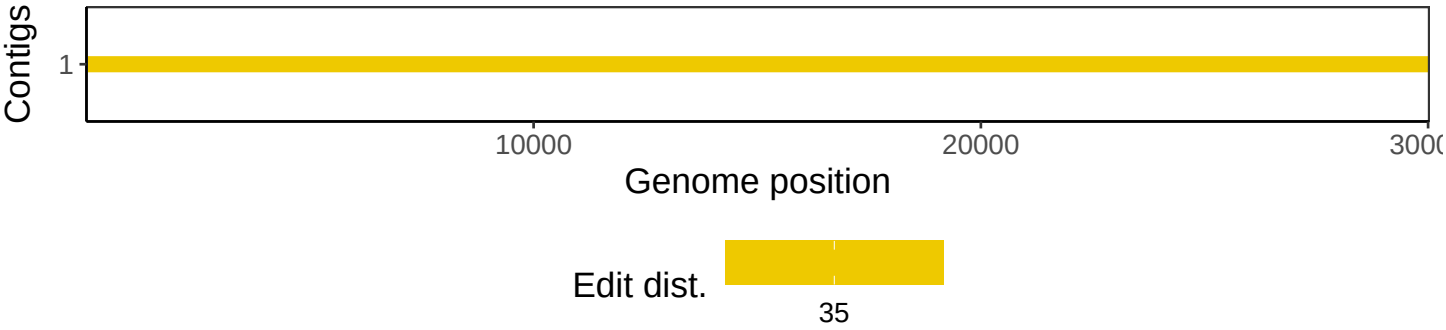
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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