COVID-19 subject UPHS-0024

2021-05-05

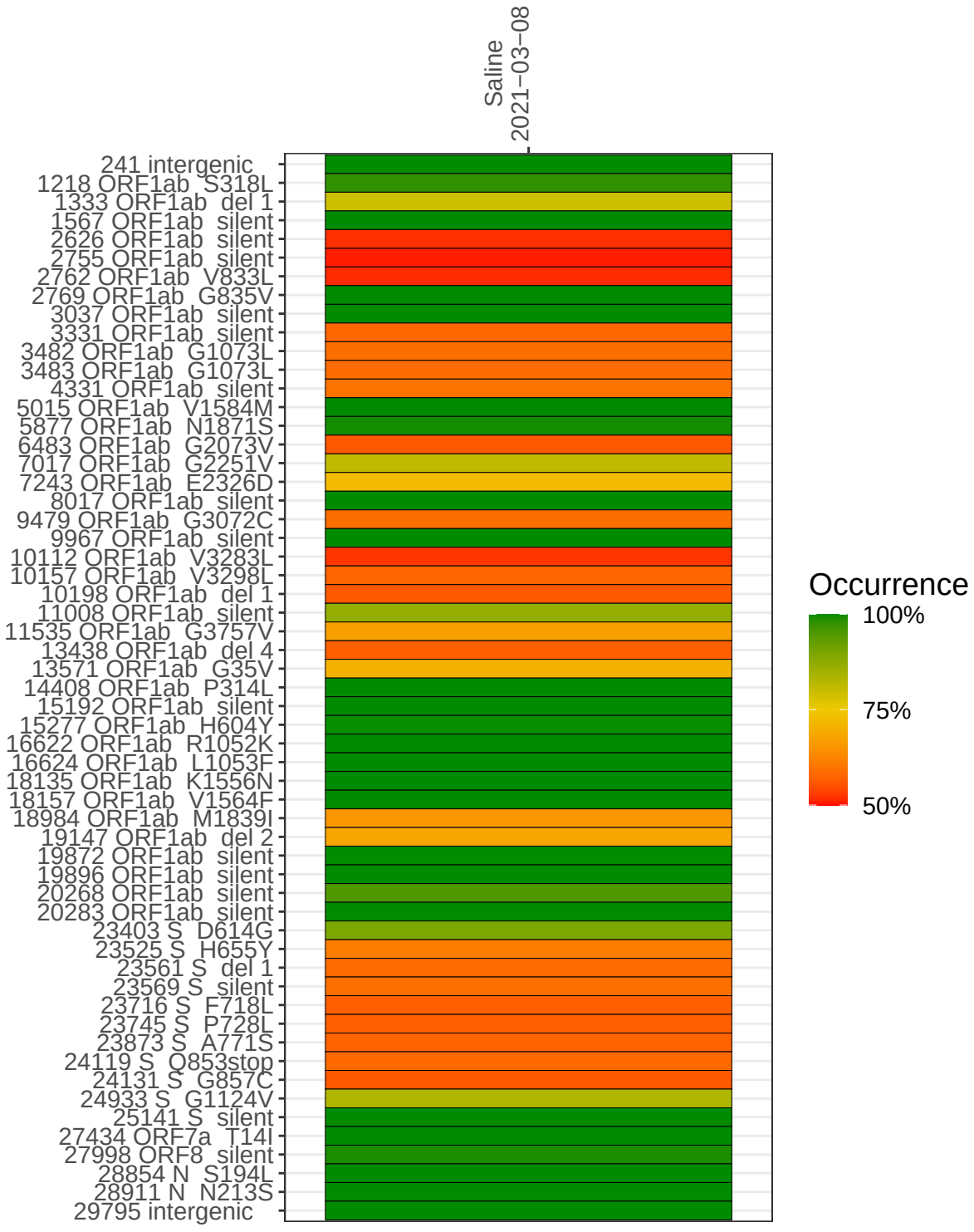
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

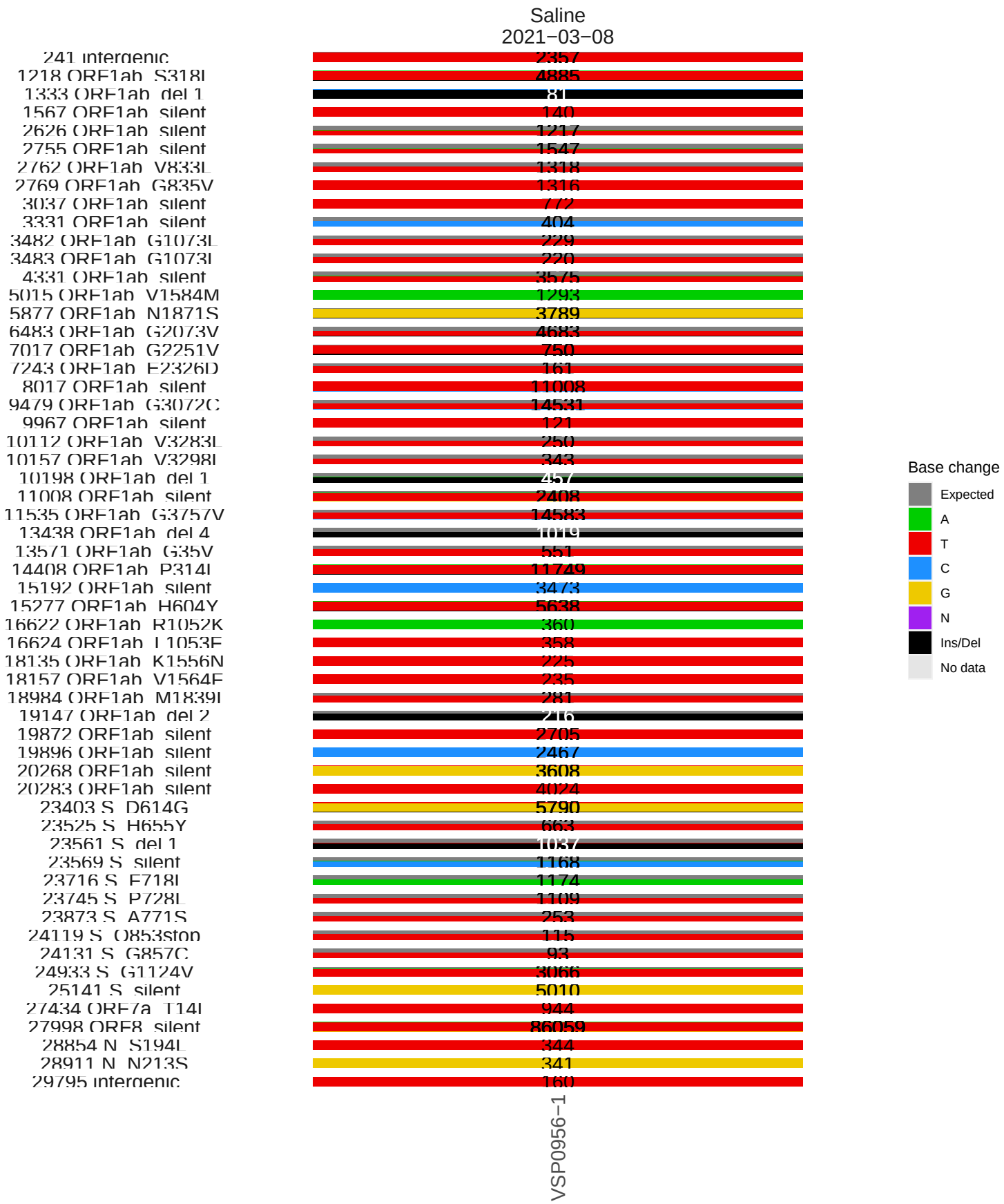
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0956-1	single experiment	NA	Saline	2021-03-08	19.49	B.1.409	97.5%	96.7%

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.

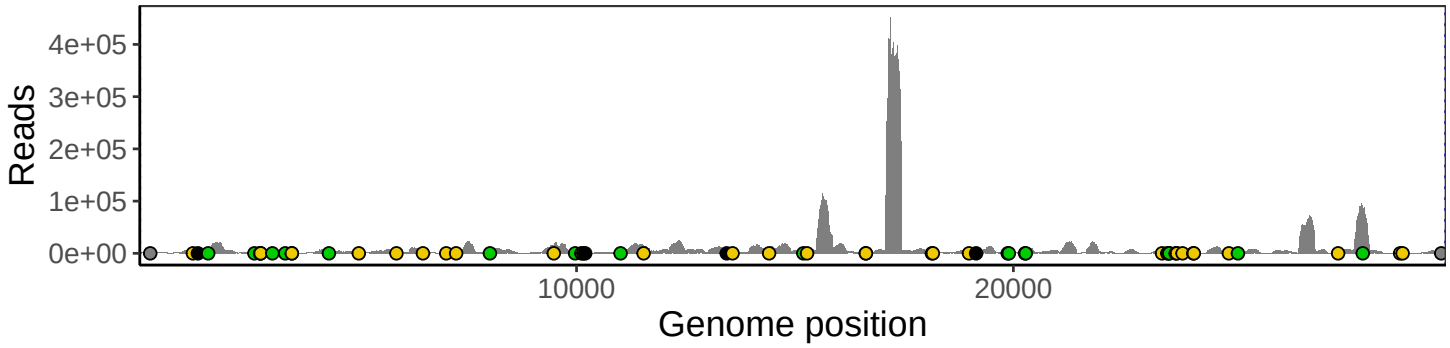



Saline
2021−03−08
241 intergenic
1218 ORF1ab S318I
1333 ORF1ab del 1
1567 ORF1ab silent
2626 ORF1ab silent
2755 ORF1ab silent
2762 ORF1ab V833L
2769 ORF1ab G835V
3037 ORF1ab silent
3331 ORF1ab silent
3482 ORF1ab G1073L
3483 ORF1ab G1073I
4331 ORF1ab silent
5015 ORF1ab V1584M
5877 ORF1ab N1871S
6483 ORF1ab G2073V
7017 ORF1ab G2251V
7243 ORF1ab E2326D
8017 ORF1ab silent
9479 ORF1ab G3072C
9967 ORF1ab silent
10112 ORF1ab V3283L
10157 ORF1ab V3298I
10198 ORF1ab del 1
11008 ORF1ab silent
11535 ORF1ab G3757V
13438 ORF1ab del 4
13571 ORF1ab G35V
14408 ORF1ab P314I
15192 ORF1ab silent
15277 ORF1ab H604Y
16622 ORF1ab R1052K
16624 ORF1ab L1053F
18135 ORF1ab K1556N
18157 ORF1ab V1564F
18984 ORF1ab M1839I
19147 ORF1ab del 2
19872 ORF1ab silent
19896 ORF1ab silent
20268 ORF1ab silent
20283 ORF1ab silent
23403 S D614G
23525 S H655Y
23561 S del 1
23569 S silent
23716 S E718I
23745 S P728L
23873 S A771S
24119 S Q853stop
24131 S G857C
24933 S G1124V
25141 S silent
27434 ORF7a T14I
27998 ORF8 silent
28854 N S194L
28911 N N213S
29795 intergenic
2357
4885
81
140
1217
1547
1318
1316
772
404
229
220
3575
1293
3789
4683
750
161
11008
14531
121
250
343
457
2408
14583
1019
551
11749
3473
5638
360
358
225
235
281
216
2705
2467
3608
4024
5790
663
1037
1168
1174
1109
253
115
93
3066
5010
944
86059
344
341
160
Base change
Expected
A
T
C
G
N
Ins/Del
No data
VSP0956−1

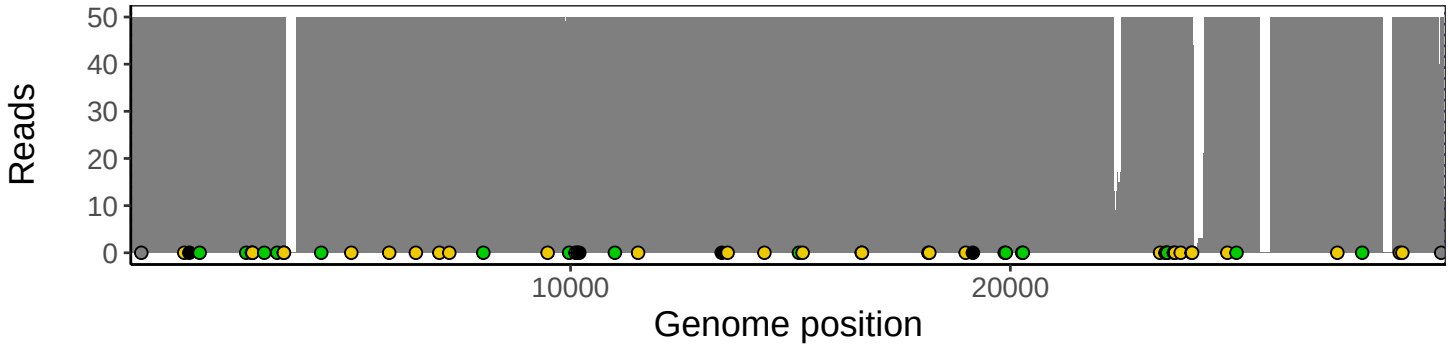
Analyses of individual experiments and composite results

VSP0956-1 | 2021-03-08 | Saline | UPHS-0024 | genomes | single experiment

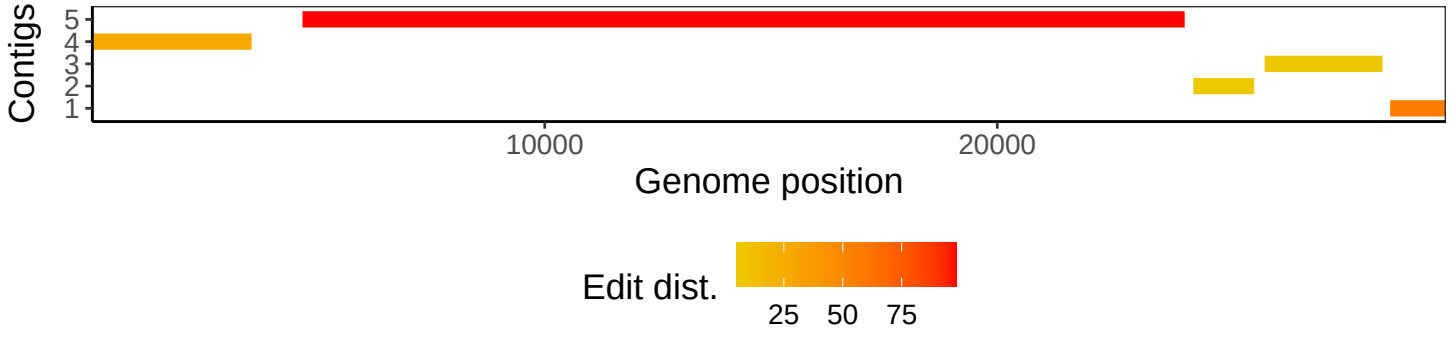
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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