COVID-19 subject UPHS-0037

2021-05-05

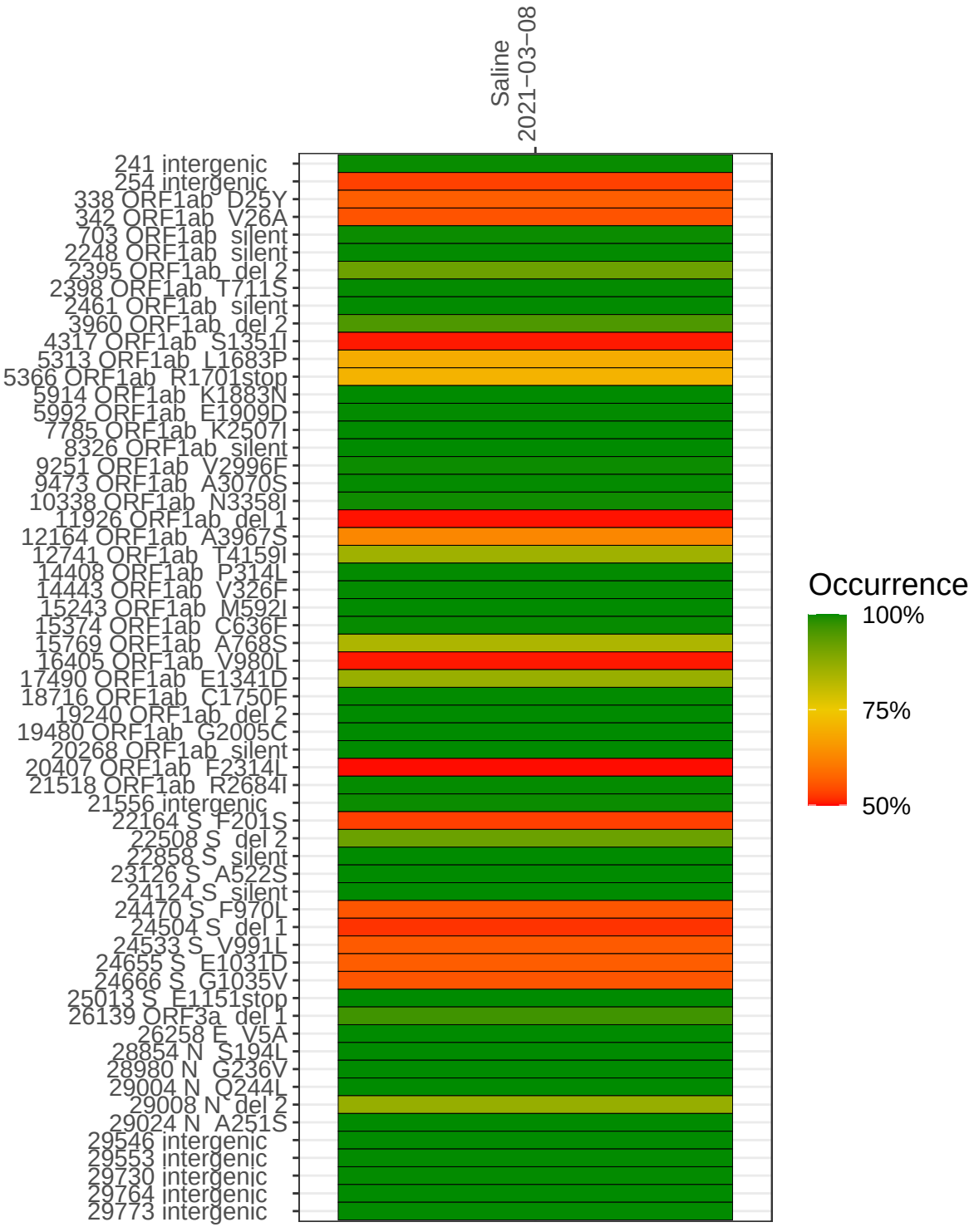
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

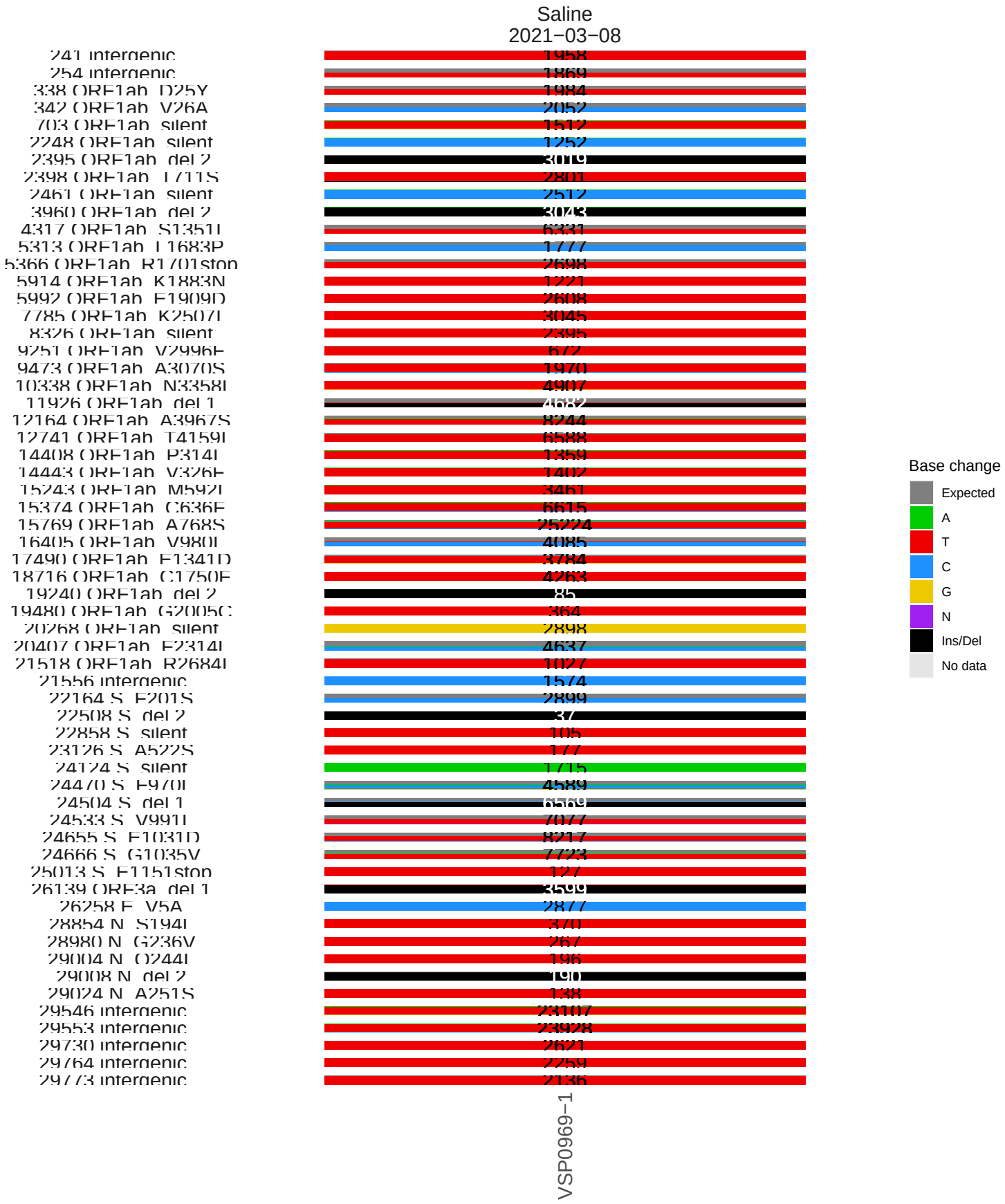
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0969-1	single experiment	NA	Saline	2021-03-08	2.48	NA	58.2%	57.2%

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.

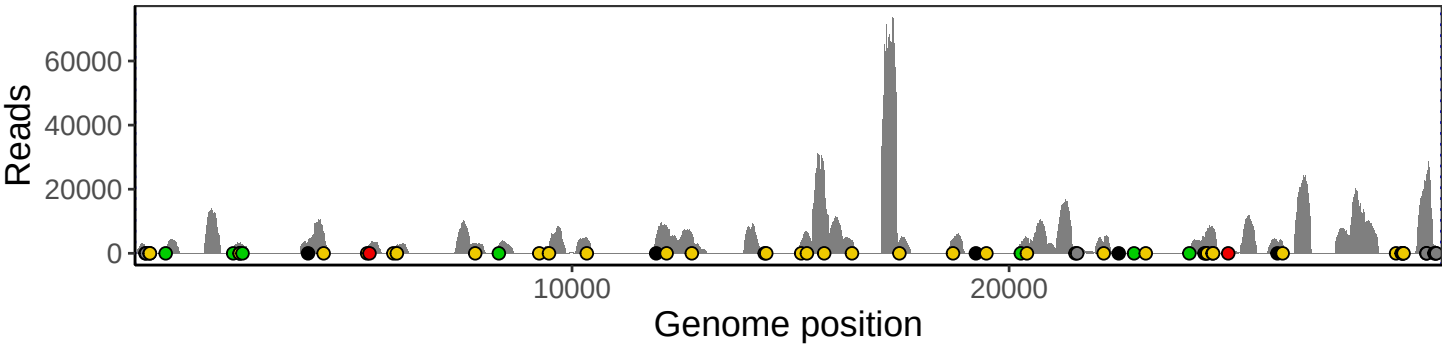



Saline
2021−03−08
241 intergenic
254 intergenic
338 ORF1ab_D25Y
342 ORF1ab_V26A
703 ORF1ab_silent
2248 ORF1ab_silent
2395 ORF1ab_del 2
2398 ORF1ab_T711S
2461 ORF1ab_silent
3960 ORF1ab_del 2
4317 ORF1ab_S1351I
5313 ORF1ab_L1683P
5366 ORF1ab_R1701stop
5914 ORF1ab_K1883N
5992 ORF1ab_F1909D
7785 ORF1ab_K2507I
8326 ORF1ab_silent
9251 ORF1ab_V2996F
9473 ORF1ab_A3070S
10338 ORF1ab_N3358I
11926 ORF1ab_del 1
12164 ORF1ab_A3967S
12741 ORF1ab_T4159I
14408 ORF1ab_P3141
14443 ORF1ab_V3256F
15243 ORF1ab_M5921
15374 ORF1ab_C6366F
15769 ORF1ab_A7685
16405 ORF1ab_V9801
17490 ORF1ab_F13411D
18716 ORF1ab_C17505F
19240 ORF1ab_del 2
19480 ORF1ab_G20050C
20268 ORF1ab_silent
20407 ORF1ab_F23141
21518 ORF1ab_R26841
21556 intergenic
22164 S_F201S
22508 S_del 2
22858 S_silent
23126 S_A522S
24124 S_silent
24470 S_F970I
24504 S_del 1
24533 S_V991I
24655 S_F1031D
24666 S_G1035V
25013 S_F1151stop
26139 ORF3a_del 1
26258 E_V5A
28854 N_S194I
28980 N_G236V
29004 N_Q244I
29008 N_del 2
29024 N_A251S
29546 intergenic
29553 intergenic
29730 intergenic
29764 intergenic
29773 intergenic
1958
1869
1984
2052
1512
1252
3019
2801
2512
3043
6331
1777
2698
1221
2608
3045
2395
672
1970
4907
4682
8244
6588
1359
1402
3461
6615
25224
4085
3784
4263
85
364
2898
4637
1027
1574
2899
37
105
177
1715
4589
6569
7077
8217
7723
127
3590
2877
370
267
196
190
138
23107
23928
2621
2259
2136
Base change
Expected
A
T
C
G
N
Ins/Del
No data
VSP0969−1

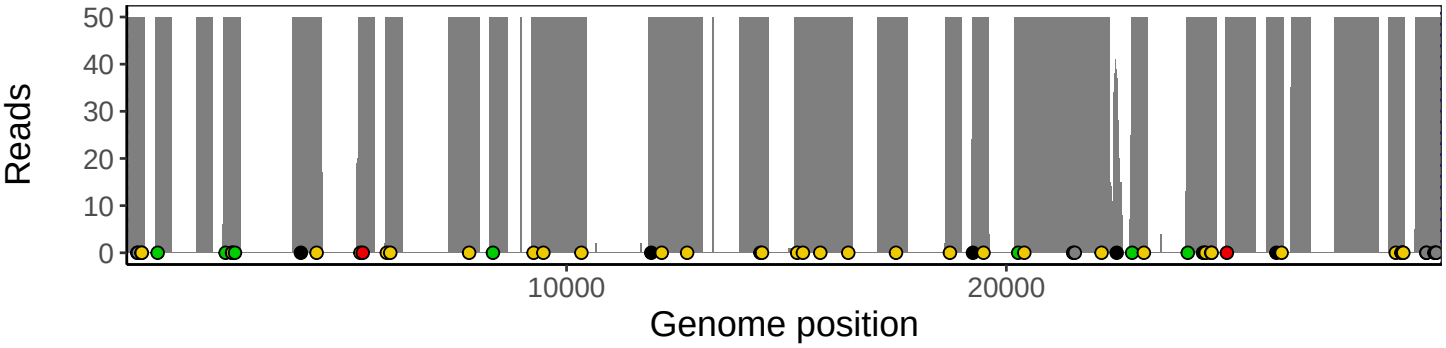
Analyses of individual experiments and composite results

VSP0969-1 | 2021-03-08 | Saline | UPHS-0037 | genomes | single experiment

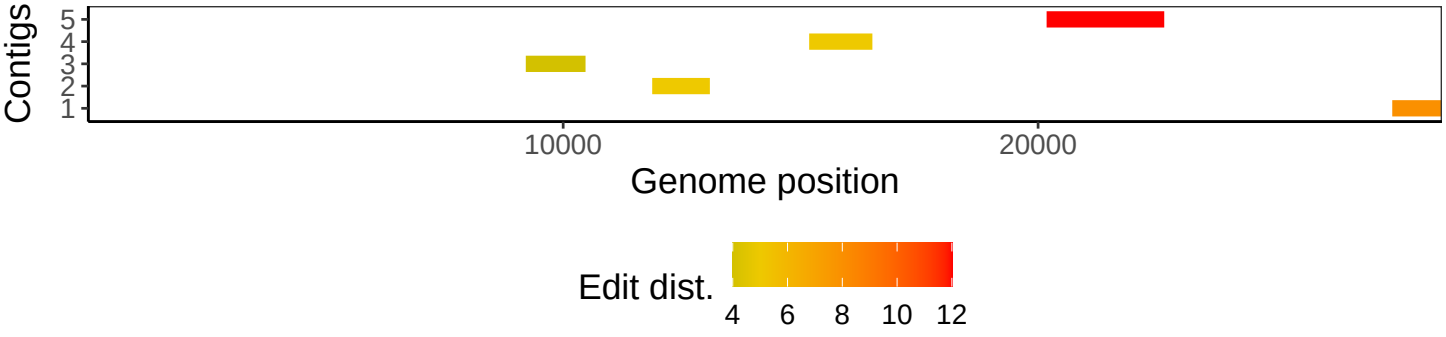
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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