COVID-19 subject HUP Q-0200

2021-06-23

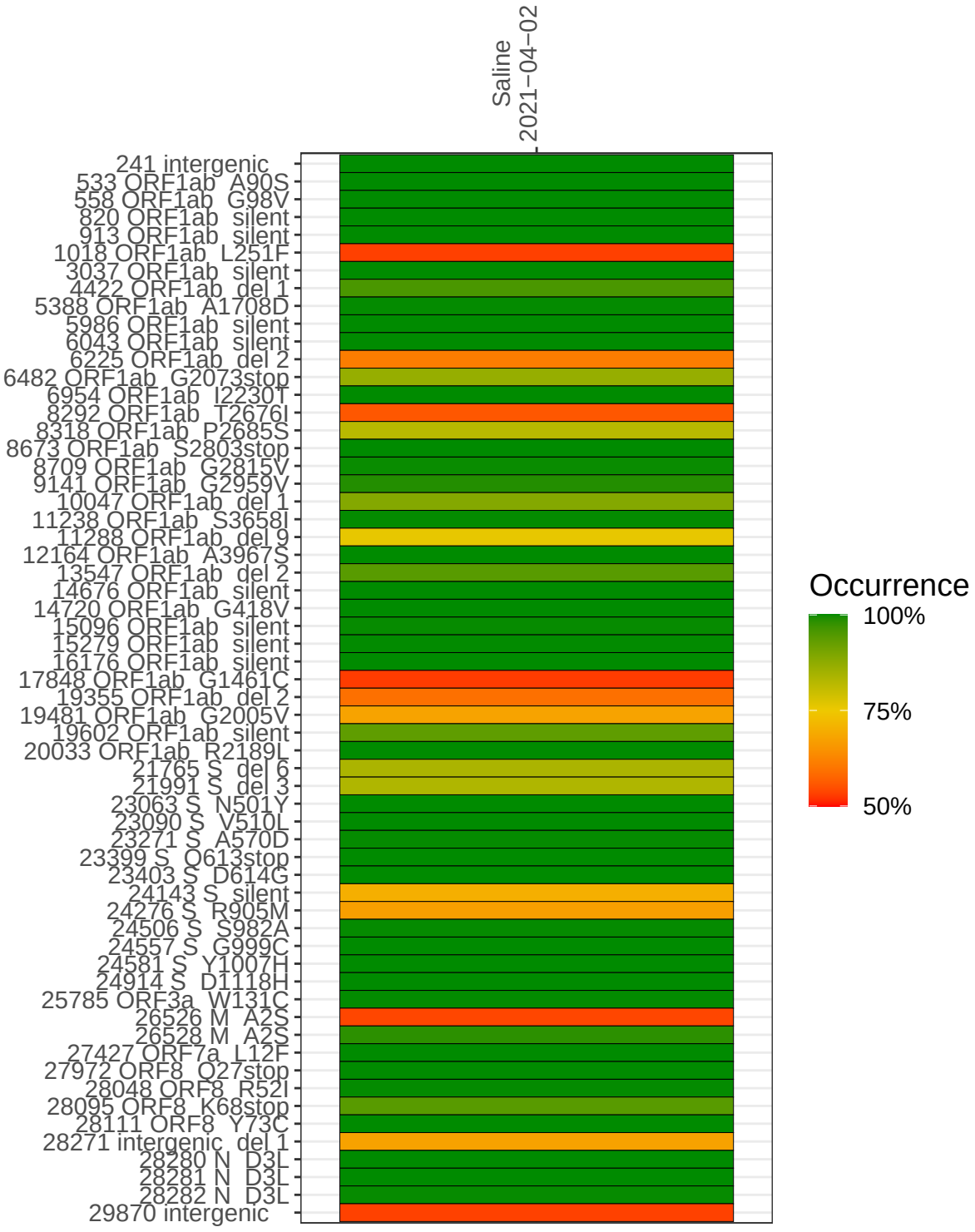
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

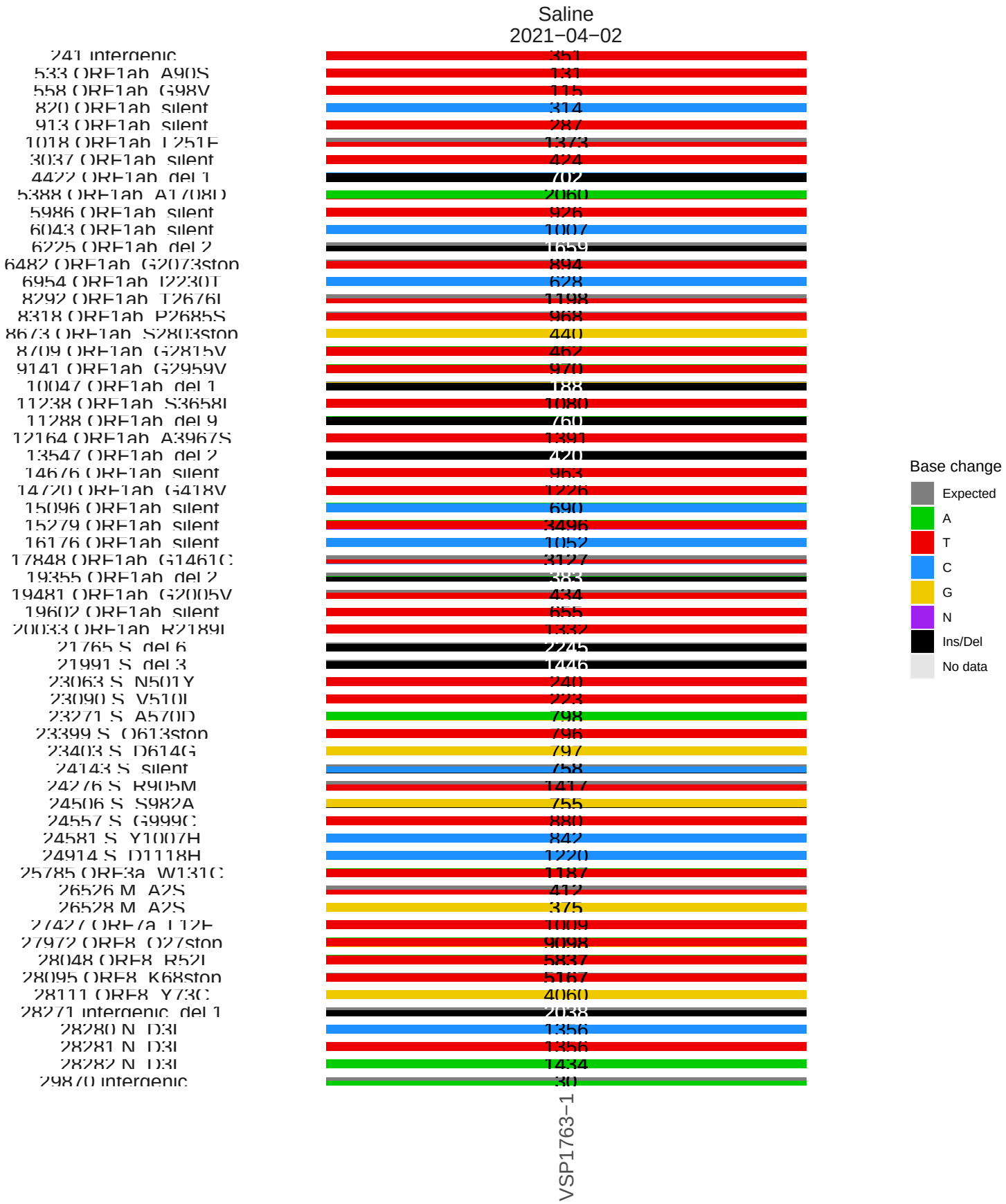
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1763-1	single experiment	NA	Saline	2021-04-02	4.39	NA	78.2%	77.9%

Variants shared across samples

The heat map below shows how variants (reference genome /home/common/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.

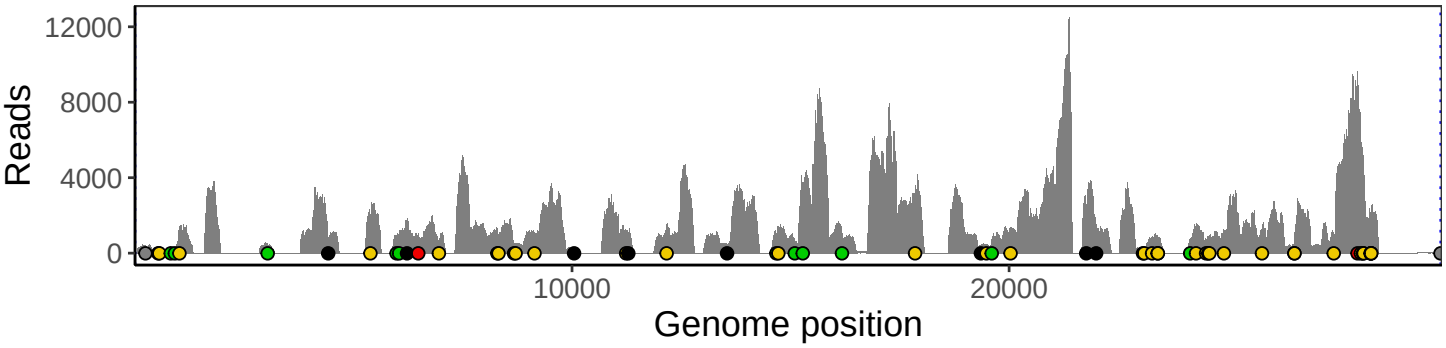



Saline
2021−04−02
241 intergenic
533 ORF1ab_A90S
558 ORF1ab_G98V
820 ORF1ab_silent
913 ORF1ab_silent
1018 ORF1ab_L251F
3037 ORF1ab_silent
4422 ORF1ab_del 1
5388 ORF1ab_A1708D
5986 ORF1ab_silent
6043 ORF1ab_silent
6225 ORF1ab_del 2
6482 ORF1ab_G2073ston
6954 ORF1ab_I2230T
8292 ORF1ab_T2676I
8318 ORF1ab_P2685S
8673 ORF1ab_S2803ston
8709 ORF1ab_G2815V
9141 ORF1ab_G2959V
10047 ORF1ab_del 1
11238 ORF1ab_S3658I
11288 ORF1ab_del 9
12164 ORF1ab_A3967S
13547 ORF1ab_del 2
14676 ORF1ab_silent
14720 ORF1ab_G4184V
15096 ORF1ab_silent
15279 ORF1ab_silent
16176 ORF1ab_silent
17848 ORF1ab_G5461C
19355 ORF1ab_del 2
19481 ORF1ab_G6005V
19602 ORF1ab_silent
20033 ORF1ab_R6189I
21765 S_del 6
21991 S_del 3
23063 S_N501Y
23090 S_V510I
23271 S_A570D
23399 S_Q613ston
23403 S_D614G
24143 S_silent
24276 S_R905M
24506 S_S982A
24557 S_G999C
24581 S_Y1007H
24914 S_D1118H
25785 ORF3a_W131C
26526 M_A2S
26528 M_A2S
27427 ORF7a_T12F
27972 ORF8_Q27ston
28048 ORF8_R52I
28095 ORF8_K68ston
28111 ORF8_Y73C
28271 intergenic_del 1
28280 N_D3I
28281 N_D3I
28282 N_D3I
29870 intergenic
351
131
115
314
287
1373
424
702
2060
926
1007
1659
894
628
1198
968
440
462
970
188
1080
760
1391
420
963
1226
690
3496
1052
3127
383
434
655
1332
2245
1446
240
223
798
796
797
758
1417
755
880
842
1220
1187
412
375
1009
9098
5837
5167
4060
2038
1356
1356
1434
30
VSP1763−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

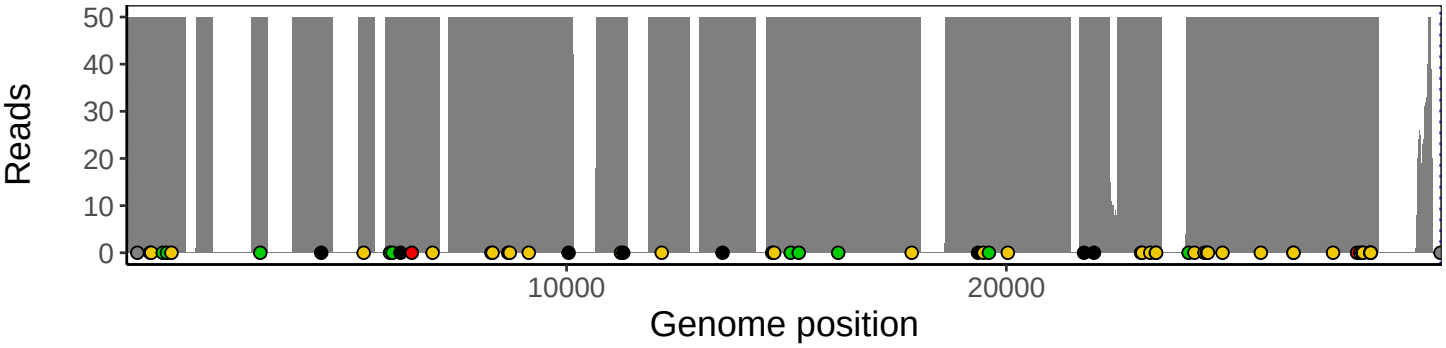
Analyses of individual experiments and composite results

VSP1763-1 | 2021-04-02 | Saline | HUP Q-0200 | genomes | single experiment

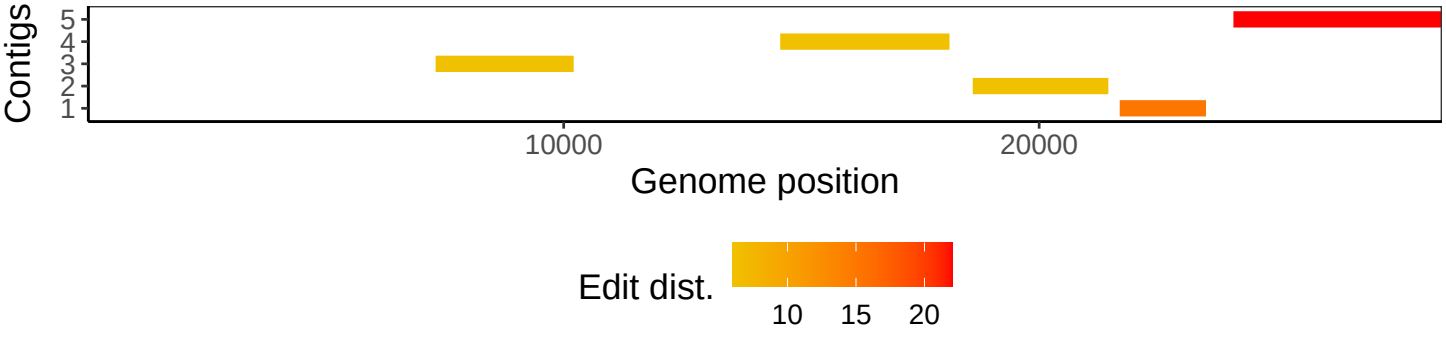
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	3.1.3
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	3.3.3
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