COVID-19 subject HUP Q-0179

2021-05-05

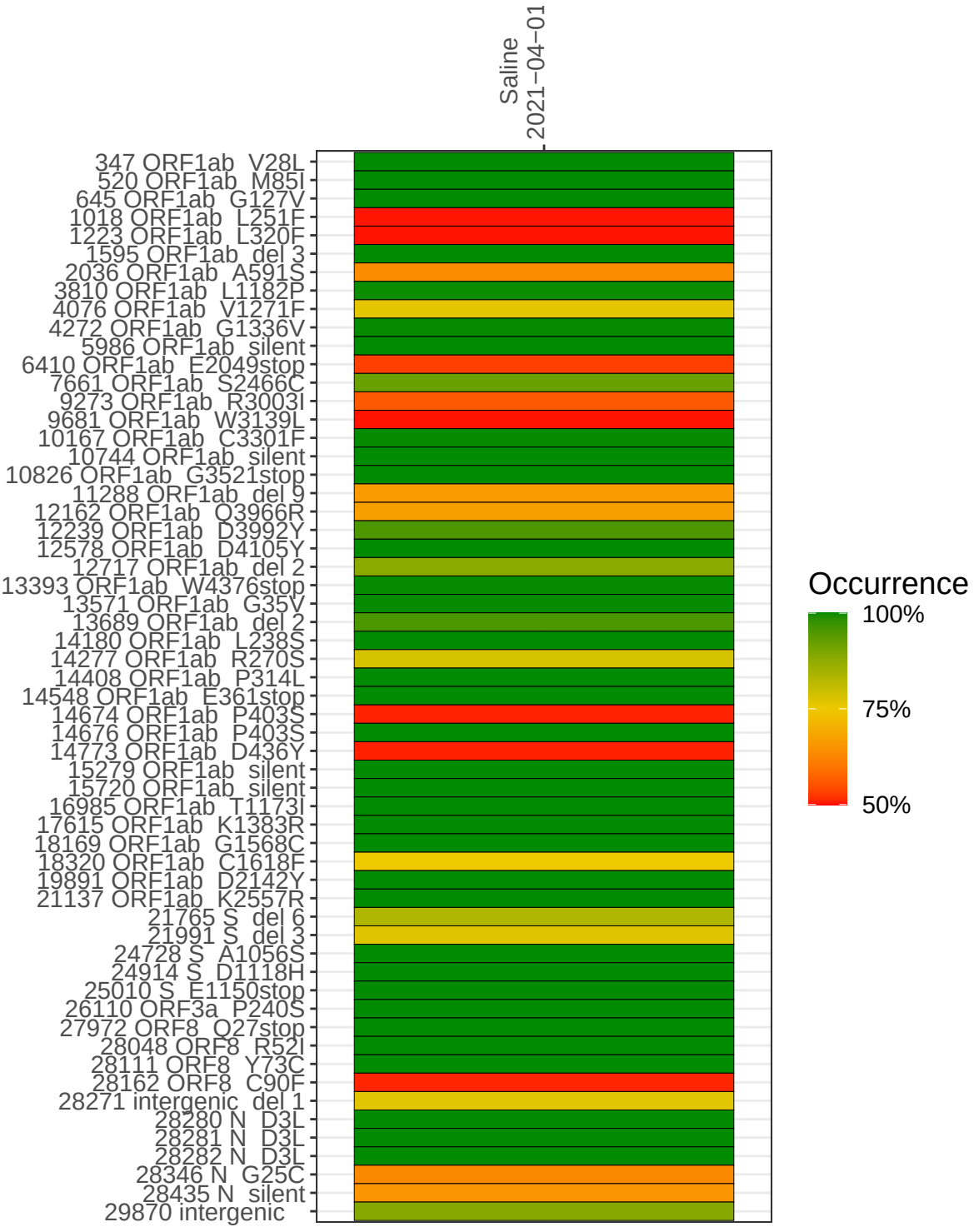
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

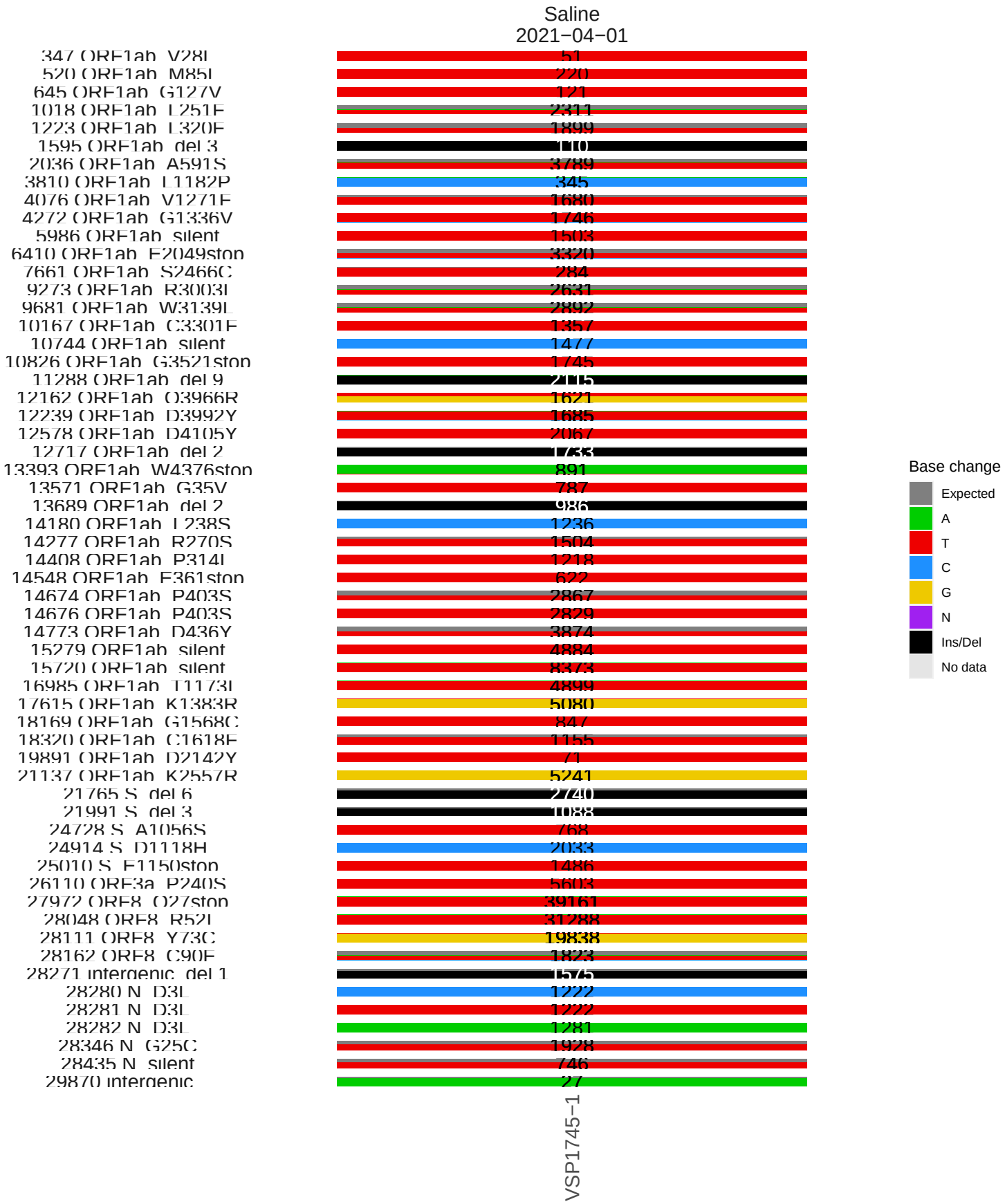
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1745-1	single experiment	NA	Saline	2021-04-01	2.61	NA	53.1%	52.7%

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.

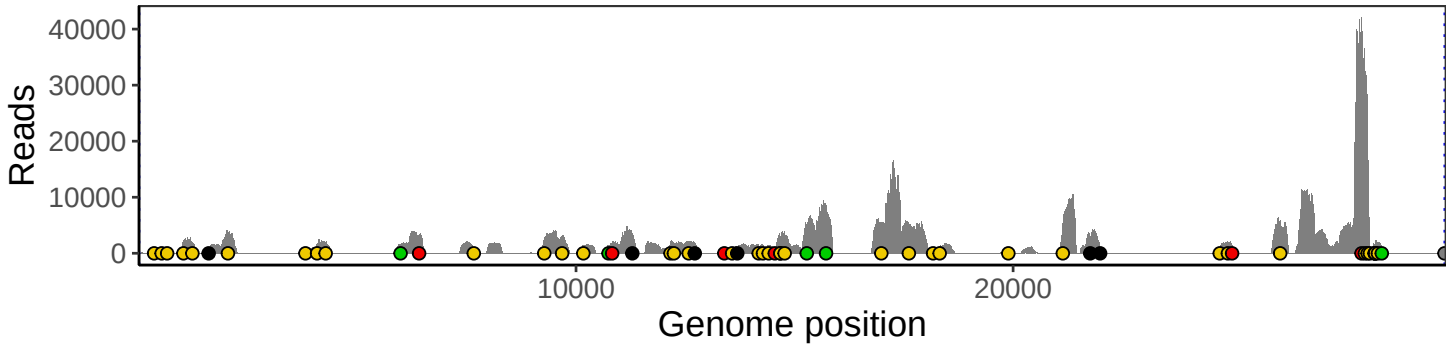



Saline
2021−04−01
347 ORF1ab V28I
520 ORF1ab M85I
645 ORF1ab G127V
1018 ORF1ab L251F
1223 ORF1ab L320F
1595 ORF1ab del 3
2036 ORF1ab A591S
3810 ORF1ab L1182P
4076 ORF1ab V1271F
4272 ORF1ab G1336V
5986 ORF1ab silent
6410 ORF1ab E2049ston
7661 ORF1ab S2466C
9273 ORF1ab R3003I
9681 ORF1ab W3139L
10167 ORF1ab C3301F
10744 ORF1ab silent
10826 ORF1ab G3521ston
11288 ORF1ab del 9
12162 ORF1ab Q3966R
12239 ORF1ab D3992Y
12578 ORF1ab D4105Y
12717 ORF1ab del 2
13393 ORF1ab W4376ston
13571 ORF1ab G35V
13689 ORF1ab del 2
14180 ORF1ab L238S
14277 ORF1ab R270S
14408 ORF1ab P314I
14548 ORF1ab E361ston
14674 ORF1ab P403S
14676 ORF1ab P403S
14773 ORF1ab D436Y
15279 ORF1ab silent
15720 ORF1ab silent
16985 ORF1ab T1173I
17615 ORF1ab K1383R
18169 ORF1ab G1568C
18320 ORF1ab C1618F
19891 ORF1ab D2142Y
21137 ORF1ab K2557R
21765 S del 6
21991 S del 3
24728 S A1056S
24914 S D1118H
25010 S E1150ston
26110 ORF3a P240S
27972 ORF8 Q27ston
28048 ORF8 R52I
28111 ORF8 Y73C
28162 ORF8 C90F
28271 intergenic del 1
28280 N D3L
28281 N D3L
28282 N D3L
28346 N G25C
28435 N silent
29870 intergenic
51
220
121
2311
1899
110
3789
345
1680
1746
1503
3320
284
2631
2892
1357
1477
1745
2115
1621
1685
2067
1733
891
787
986
1236
1504
1218
622
2867
2829
3874
4884
8373
4899
5080
847
1155
71
5241
2740
1088
768
2033
1486
5603
39161
31288
19838
1823
1575
1222
1222
1281
1928
746
27
VSP1745−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

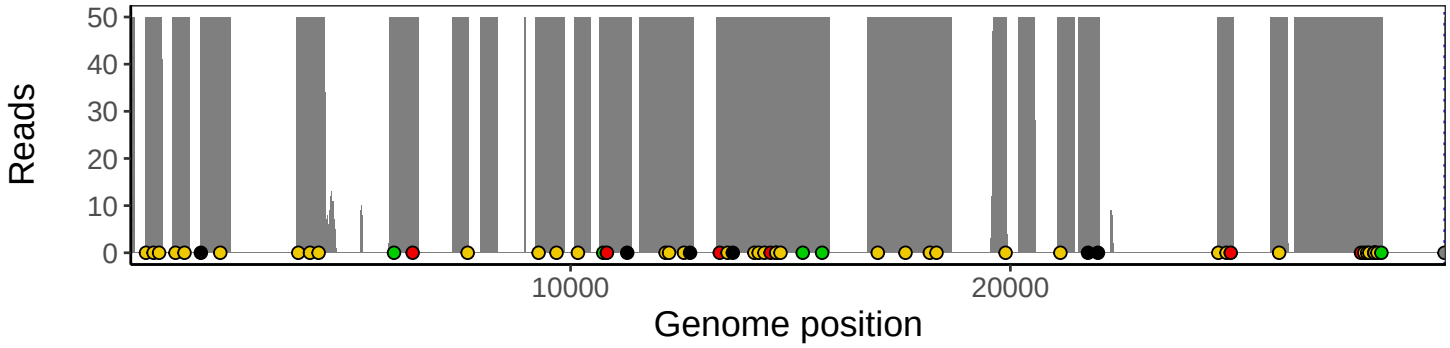
Analyses of individual experiments and composite results

VSP1745-1 | 2021-04-01 | Saline | HUP Q-0179 | genomes | single experiment

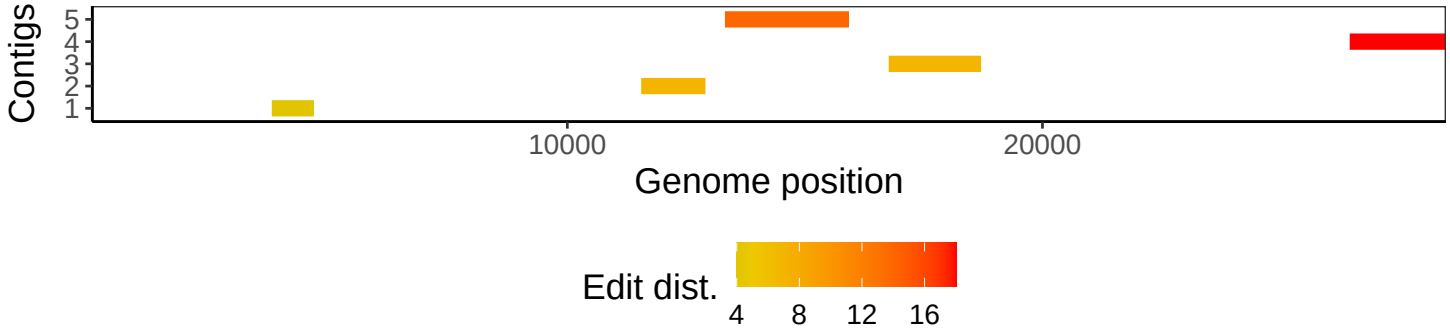
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	3.3.3
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