COVID-19 subject HUP Q-0209

2021-05-05

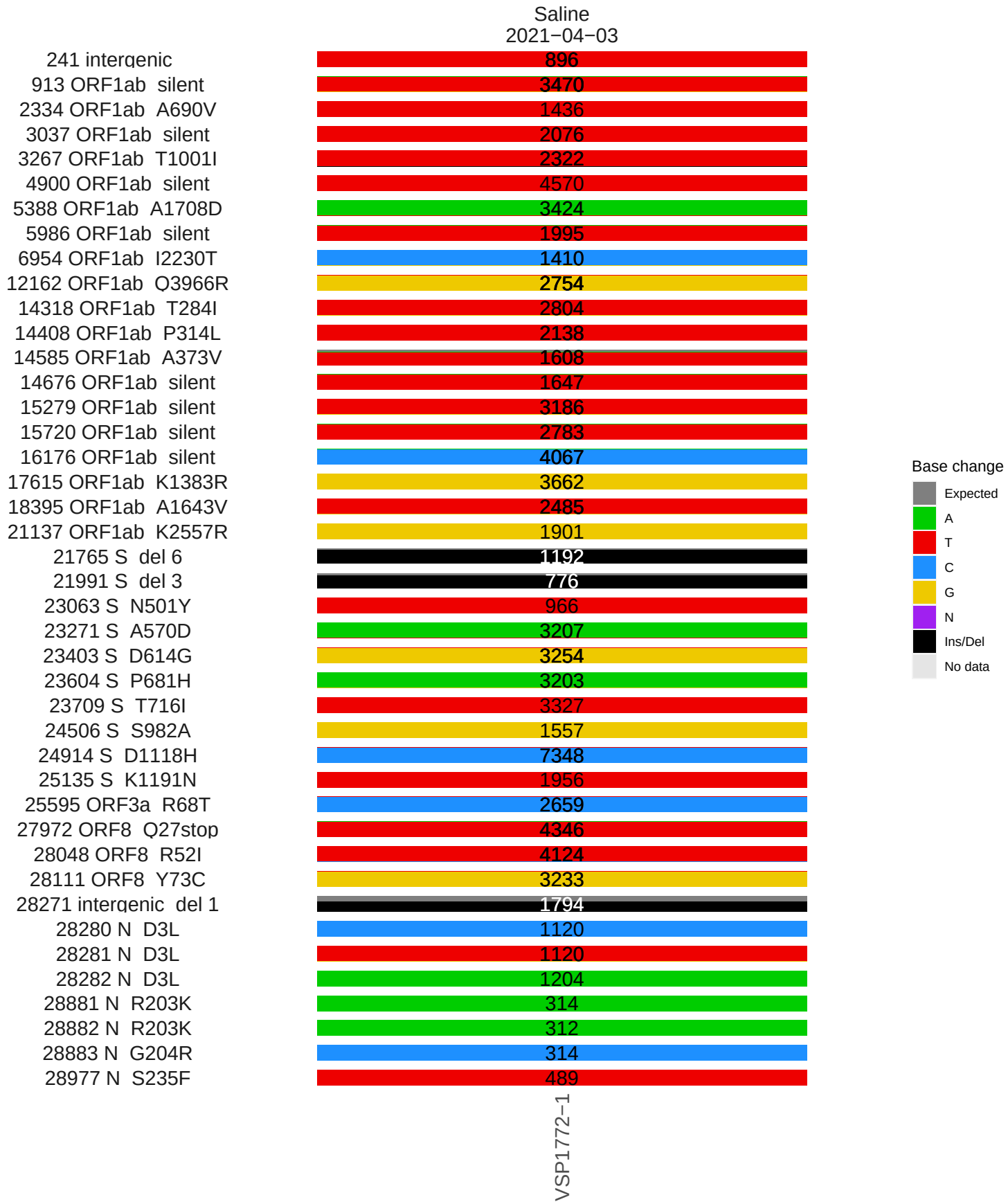
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1772-1	single experiment	NA	Saline	2021-04-03	29.86	B.1.1.7	99.8%	99.7%

Variants shared across samples

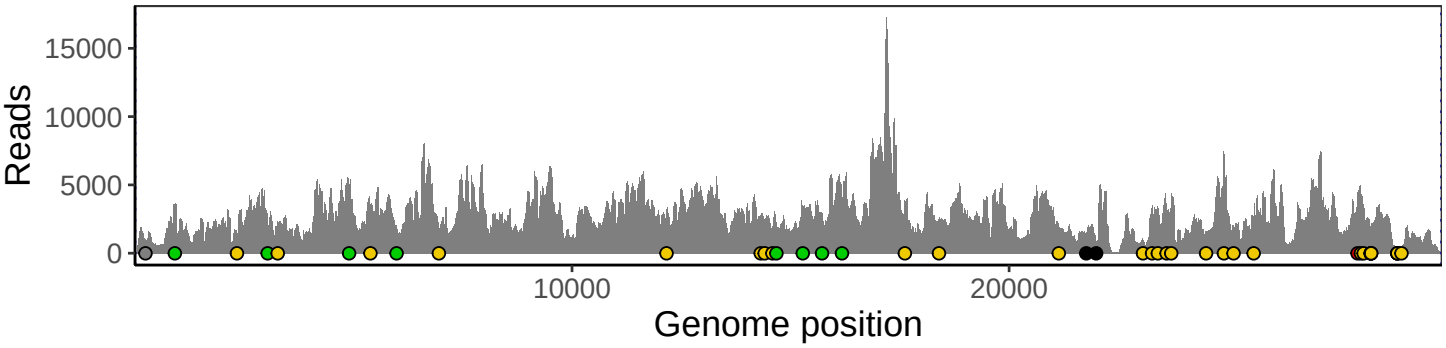
The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/Wuhan-Hu-1) are shared across subject samples where the percent variance is colored. Variants are called if a variant position is covered by 5 or more reads, the alternative base is found in > 50% of 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
2021−04−03
241 intergenic 896
913 ORF1ab silent 3470
2334 ORF1ab A690V 1436
3037 ORF1ab silent 2076
3267 ORF1ab T1001I 2322
4900 ORF1ab silent 4570
5388 ORF1ab A1708D 3424
5986 ORF1ab silent 1995
6954 ORF1ab I2230T 1410
12162 ORF1ab Q3966R 2754
14318 ORF1ab T284I 2804
14408 ORF1ab P314L 2138
14585 ORF1ab A373V 1608
14676 ORF1ab silent 1647
15279 ORF1ab silent 3186
15720 ORF1ab silent 2783
16176 ORF1ab silent 4067
17615 ORF1ab K1383R 3662
18395 ORF1ab A1643V 2485
21137 ORF1ab K2557R 1901
21765 S del 6 1192
21991 S del 3 776
23063 S N501Y 966
23271 S A570D 3207
23403 S D614G 3254
23604 S P681H 3203
23709 S T716I 3327
24506 S S982A 1557
24914 S D1118H 7348
25135 S K1191N 1956
25595 ORF3a R68T 2659
27972 ORF8 Q27stop 4346
28048 ORF8 R52I 4124
28111 ORF8 Y73C 3233
28271 intergenic del 1 1794
28280 N D3L 1120
28281 N D3L 1120
28282 N D3L 1204
28881 N R203K 314
28882 N R203K 312
28883 N G204R 314
28977 N S235F 489
VSP1772−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

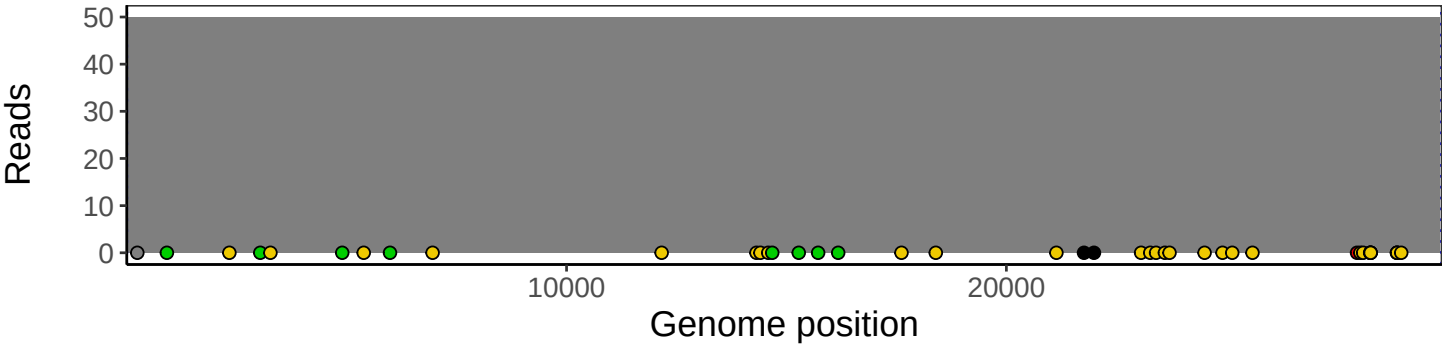
Analyses of individual experiments and composite results

VSP1772-1 | 2021-04-03 | Saline | HUP Q-0209 | genomes | single experiment

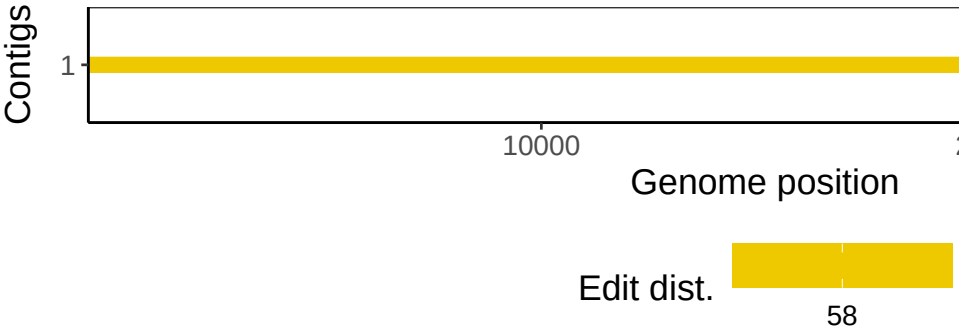
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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