COVID-19 subject HUP Q-0032

2021-04-17

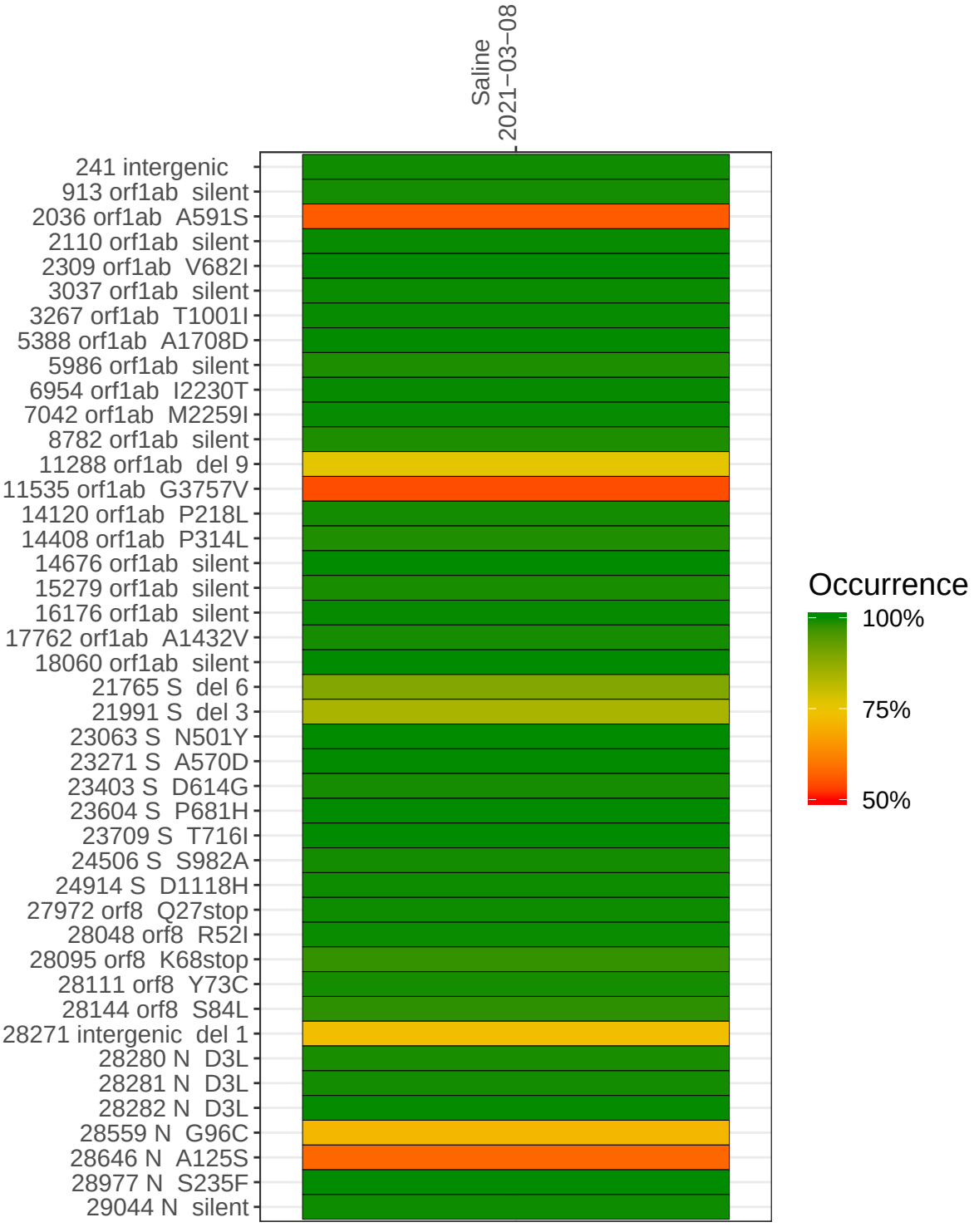
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

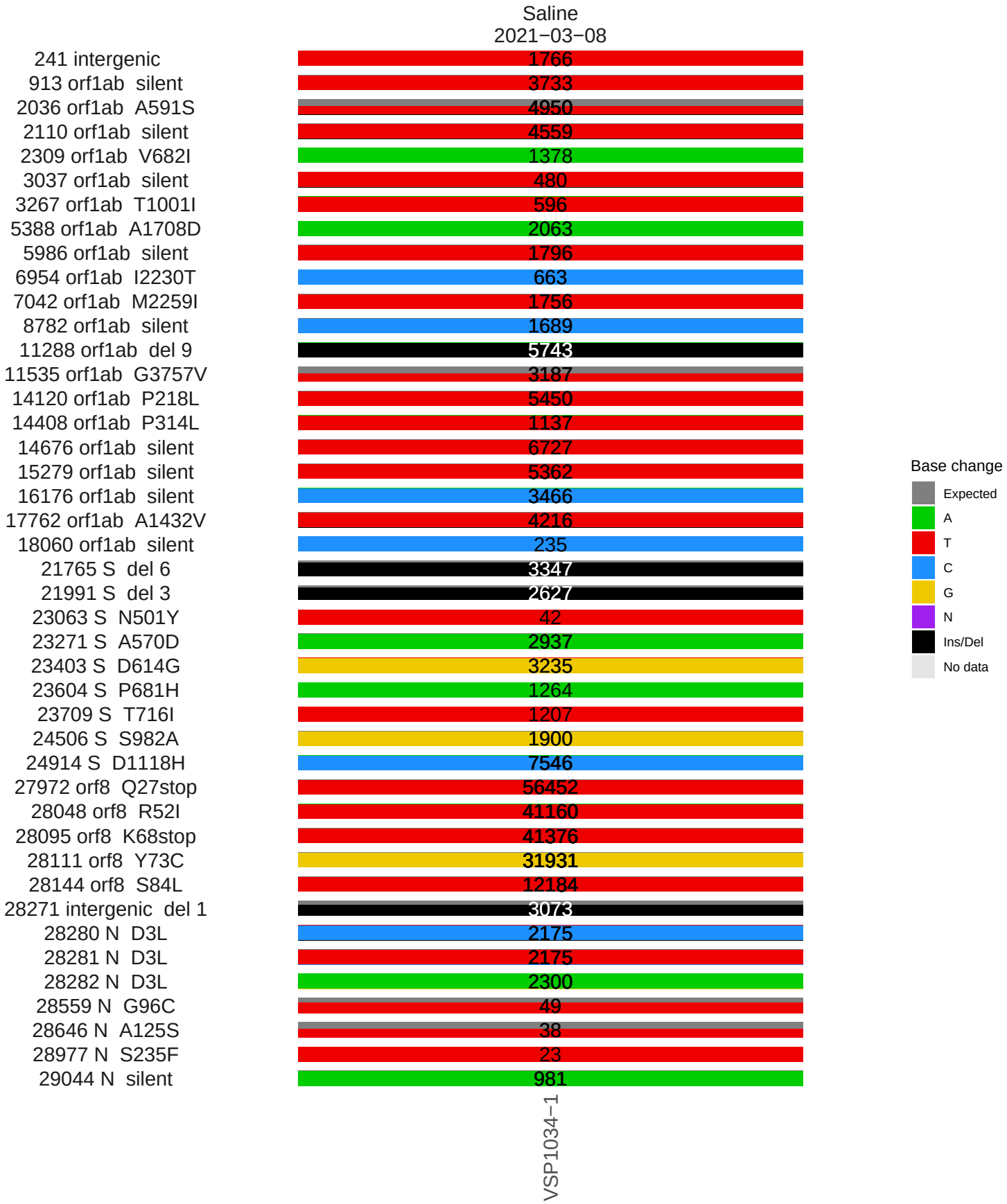
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1034-1	single experiment	NA	Saline	2021-03-08	29.83	B.1.1.7	99.8%	99.8%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/USA-WA1-2020) 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.

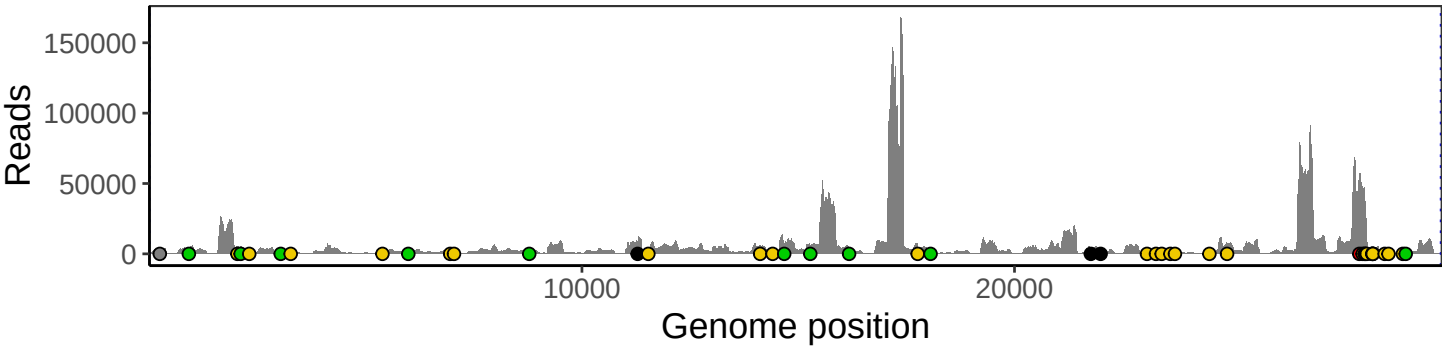



Saline
2021−03−08
241 intergenic 1766
913 orf1ab silent 3733
2036 orf1ab A591S 4950
2110 orf1ab silent 4559
2309 orf1ab V682I 1378
3037 orf1ab silent 480
3267 orf1ab T1001I 596
5388 orf1ab A1708D 2063
5986 orf1ab silent 1796
6954 orf1ab I2230T 663
7042 orf1ab M2259I 1756
8782 orf1ab silent 1689
11288 orf1ab del 9 5743
11535 orf1ab G3757V 3187
14120 orf1ab P218L 5450
14408 orf1ab P314L 1137
14676 orf1ab silent 6727
15279 orf1ab silent 5362
16176 orf1ab silent 3466
17762 orf1ab A1432V 4216
18060 orf1ab silent 235
21765 S del 6 3347
21991 S del 3 2627
23063 S N501Y 42
23271 S A570D 2937
23403 S D614G 3235
23604 S P681H 1264
23709 S T716I 1207
24506 S S982A 1900
24914 S D1118H 7546
27972 orf8 Q27stop 56452
28048 orf8 R52I 41160
28095 orf8 K68stop 41376
28111 orf8 Y73C 31931
28144 orf8 S84L 12184
28271 intergenic del 1 3073
28280 N D3L 2175
28281 N D3L 2175
28282 N D3L 2300
28559 N G96C 49
28646 N A125S 38
28977 N S235F 23
29044 N silent 981
VSP1034−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

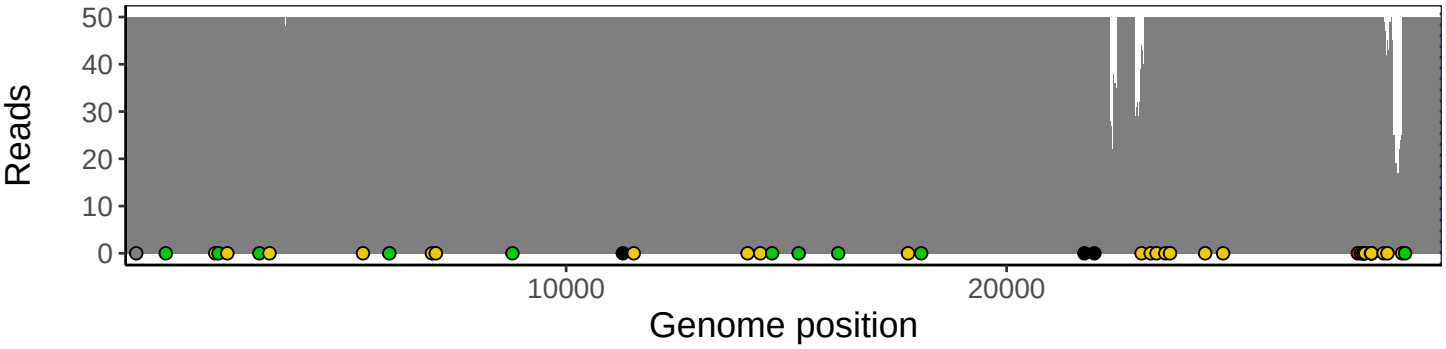
Analyses of individual experiments and composite results

VSP1034-1 | 2021-03-08 | Saline | HUP Q-0032 | genomes | single experiment

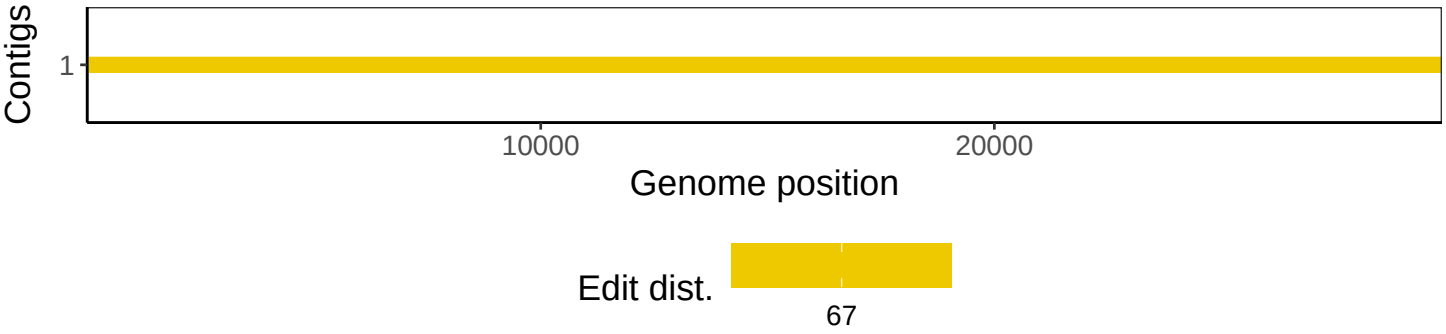
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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