COVID-19 subject HUP Q-0030

2021-04-17

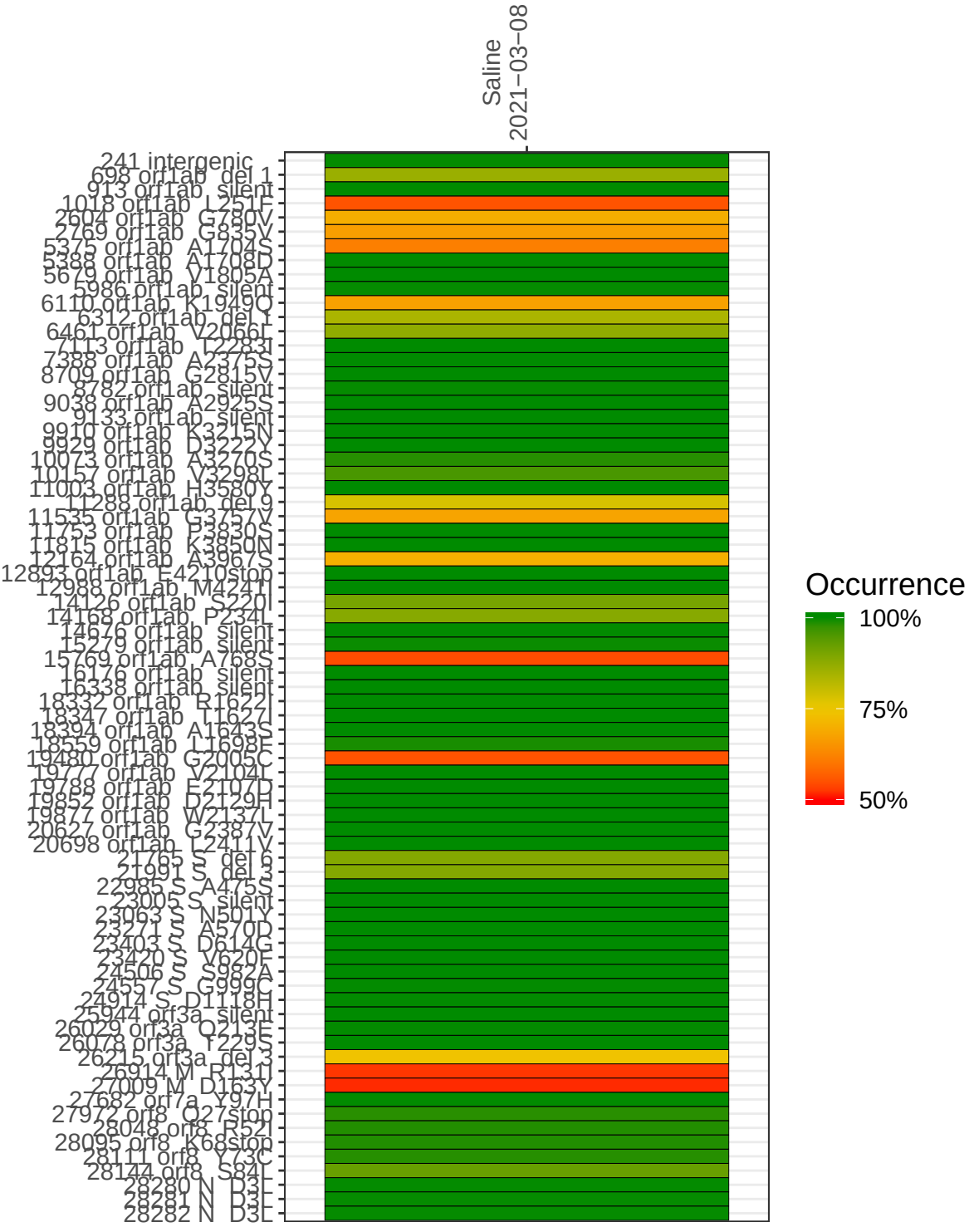
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

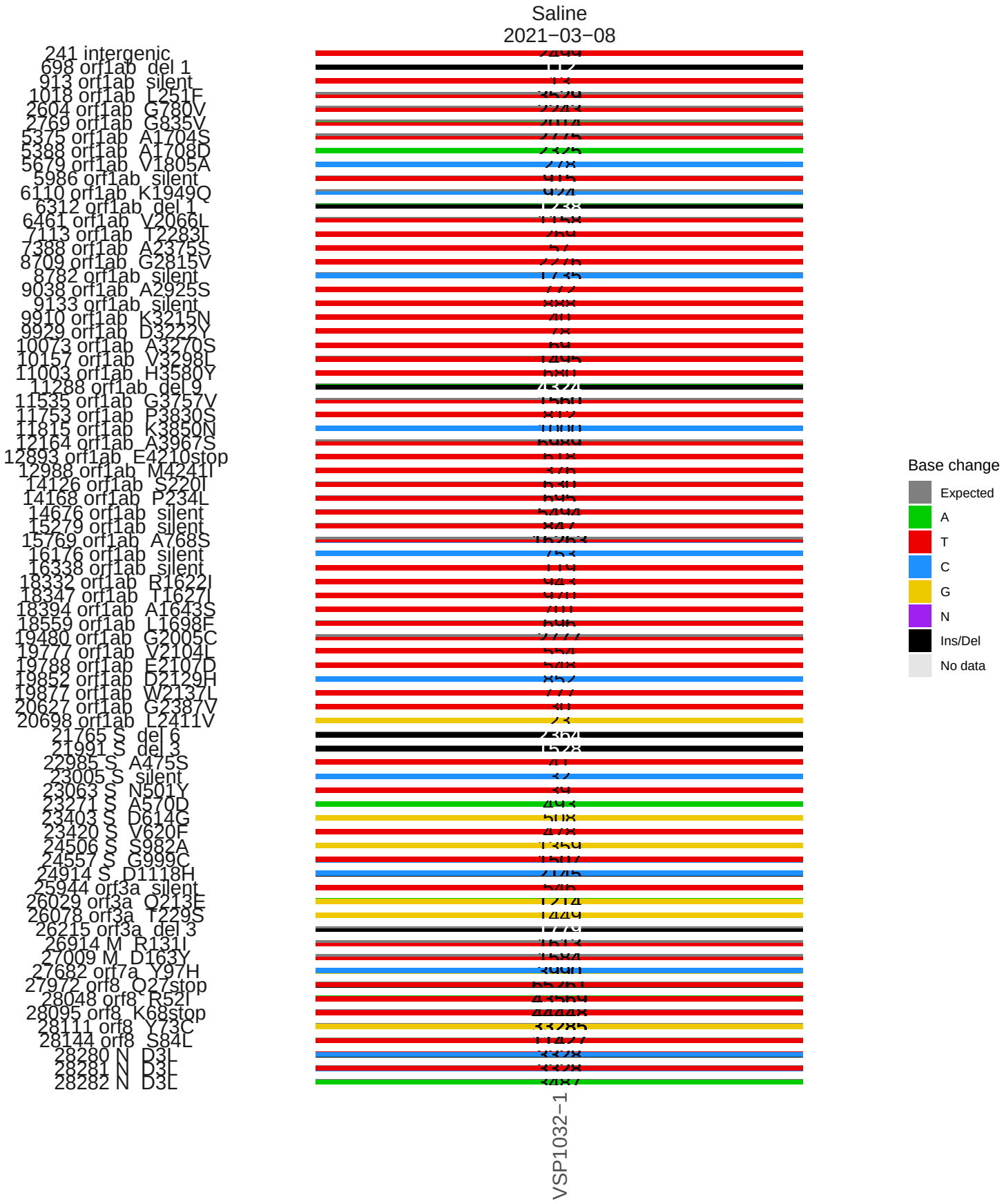
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1032-1	single experiment	NA	Saline	2021-03-08	2.82	NA	74.3%	73.4%

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.

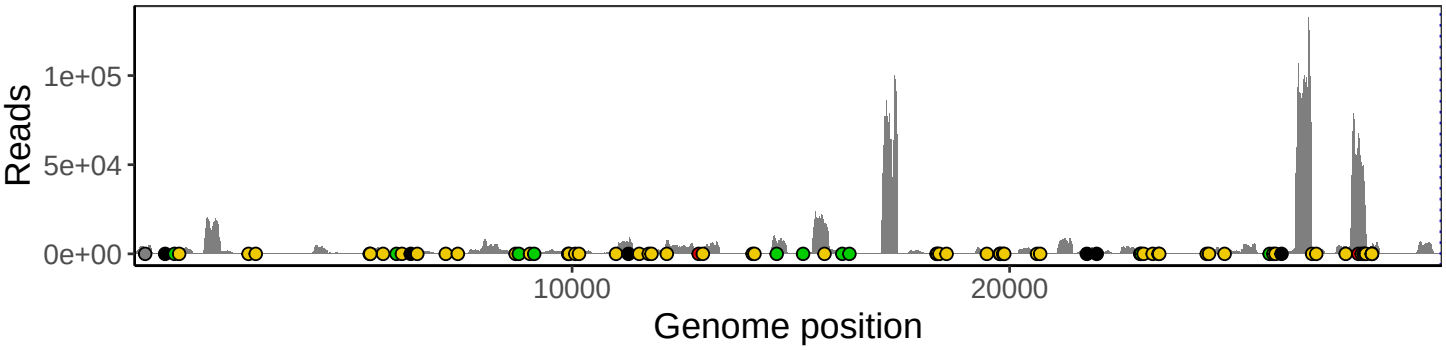



Saline
2021−03−08
241 intergenic
698 orf1ab del 1
913 orf1ab silent
1018 orf1ab L251F
2604 orf1ab G780V
2769 orf1ab G835V
5375 orf1ab A1704S
5388 orf1ab A1708D
5679 orf1ab V1805A
5986 orf1ab silent
6110 orf1ab K1949Q
6312 orf1ab del 1
6461 orf1ab V2066L
7113 orf1ab T2283I
7388 orf1ab A2375S
8709 orf1ab G2815V
8782 orf1ab silent
9038 orf1ab A2925S
9133 orf1ab silent
9910 orf1ab K3215N
9929 orf1ab D3222Y
10073 orf1ab A3270S
10157 orf1ab V3298L
11003 orf1ab H3580Y
11288 orf1ab del 9
11535 orf1ab G3757V
11753 orf1ab P3830S
11815 orf1ab K3850N
12164 orf1ab A3967S
12893 orf1ab E4210stop
12988 orf1ab M4241I
14126 orf1ab S220I
14168 orf1ab P234L
14676 orf1ab silent
15279 orf1ab silent
15769 orf1ab A768S
16176 orf1ab silent
16338 orf1ab silent
18332 orf1ab R1622I
18347 orf1ab T1627I
18394 orf1ab A1643S
18559 orf1ab L1698F
19480 orf1ab G2005C
19777 orf1ab V2104L
19788 orf1ab E2107D
19852 orf1ab D2129H
19877 orf1ab W2137L
20627 orf1ab G2387V
20698 orf1ab L2411V
21765 S del 6
21991 S del 3
22985 S A475S
23005 S silent
23063 S N501Y
23271 S A570D
23403 S D614G
23420 S V620F
24506 S S982A
24557 S G999C
24914 S D1118H
25944 orf3a silent
26029 orf3a Q213E
26078 orf3a T229S
26215 orf3a del 3
26914 M R131I
27009 M D163Y
27682 orf7a Y97H
27972 orf8 Q27stop
28048 orf8 R52I
28095 orf8 K68stop
28111 orf8 Y73C
28144 orf8 S84L
28280 N D3L
28281 N D3L
28282 N D3L
Base change
Expected
A
T
C
G
N
Ins/Del
No data
VSP1032−1

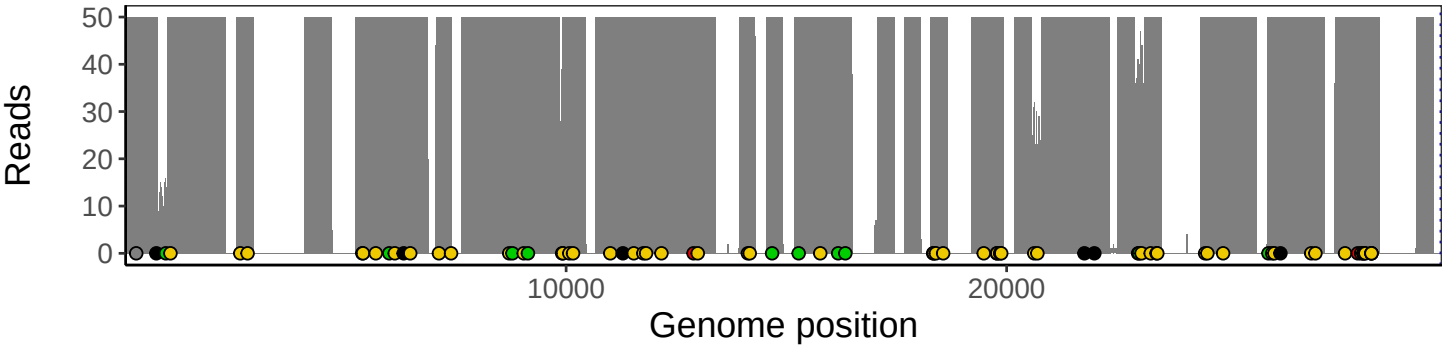
Analyses of individual experiments and composite results

VSP1032-1 | 2021-03-08 | Saline | HUP Q-0030 | genomes | single experiment

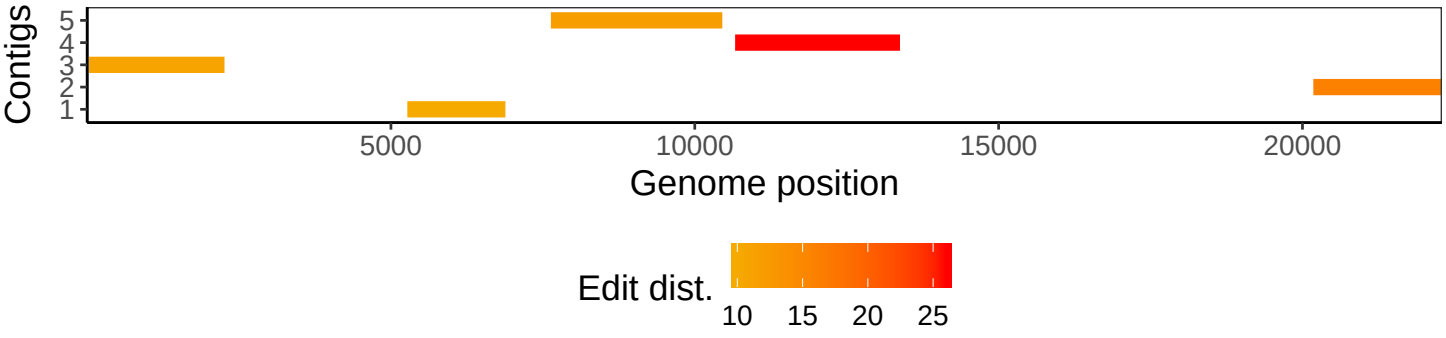
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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