COVID-19 subject H2103080875

2021-06-23

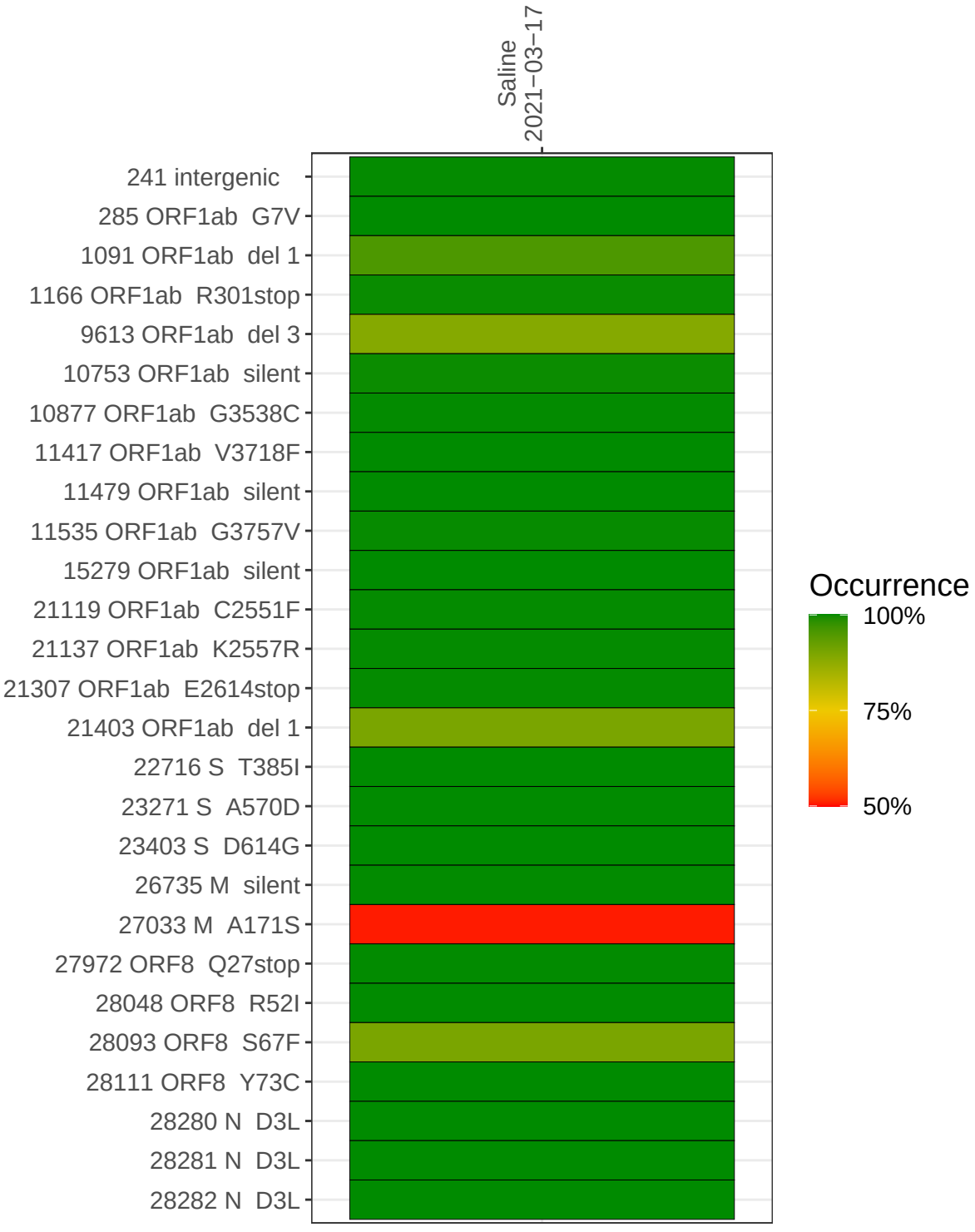
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

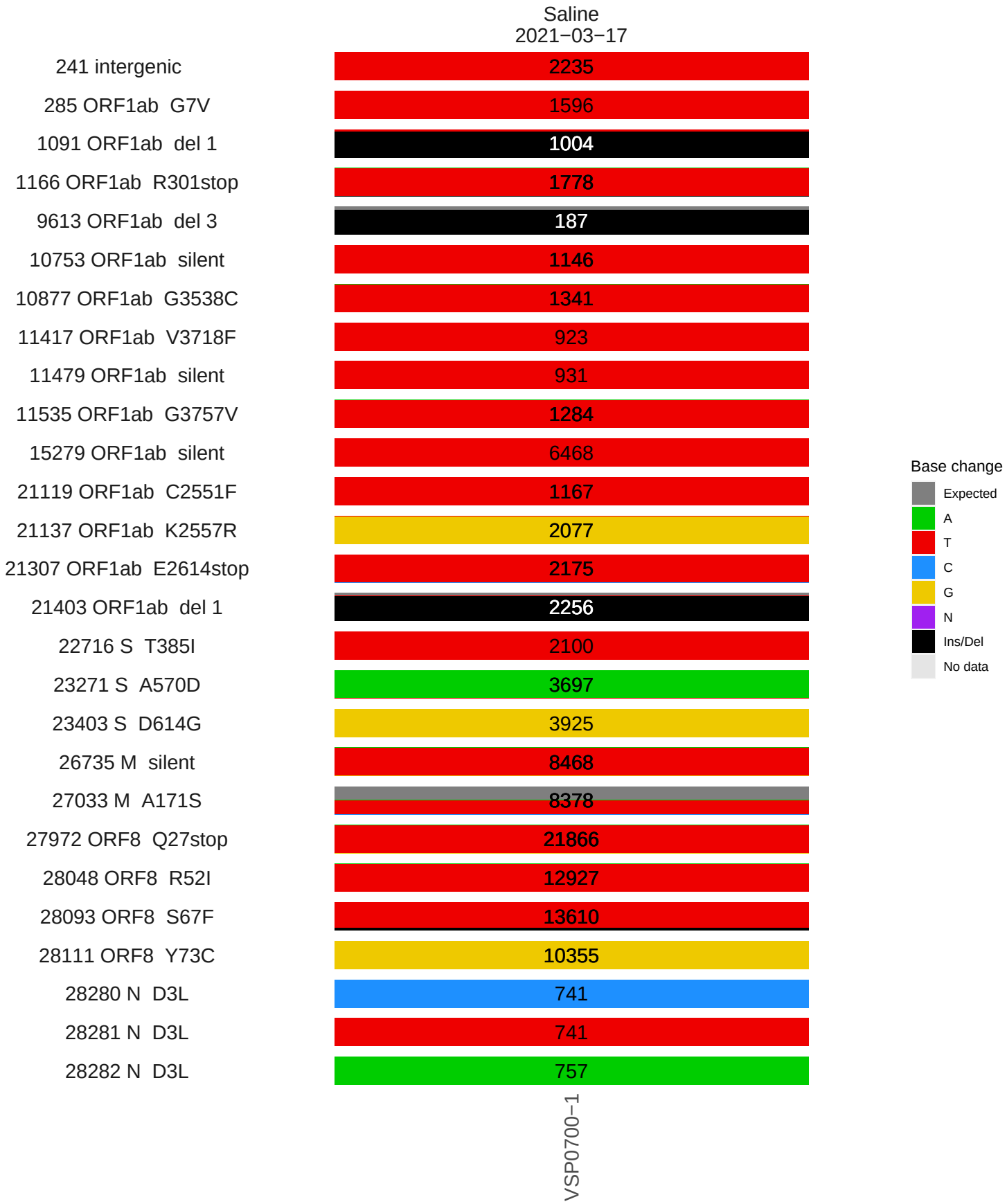
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0700-1	single experiment	NA	Saline	2021-03-17	0.75	NA	21.9%	21.0%

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.

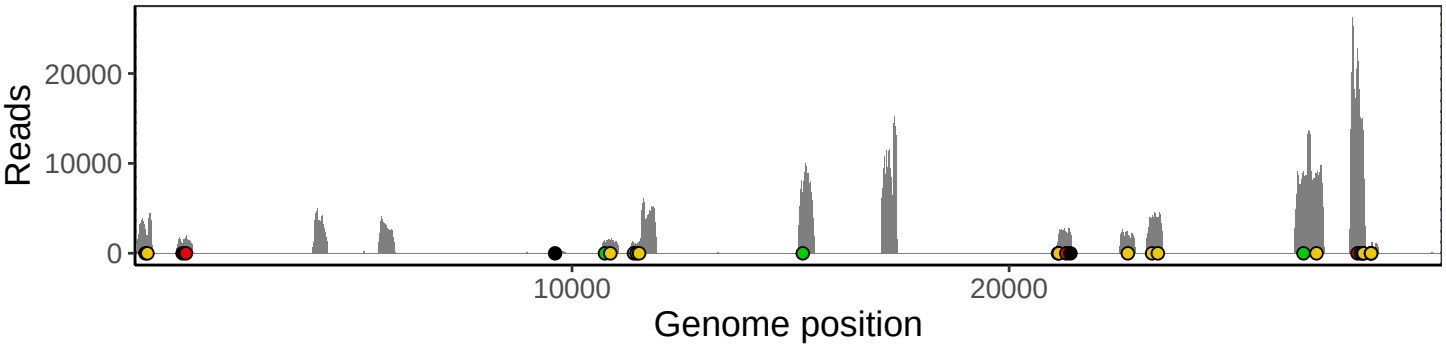



Saline
2021−03−17
241 intergenic
285 ORF1ab G7V
1091 ORF1ab del 1
1166 ORF1ab R301stop
9613 ORF1ab del 3
10753 ORF1ab silent
10877 ORF1ab G3538C
11417 ORF1ab V3718F
11479 ORF1ab silent
11535 ORF1ab G3757V
15279 ORF1ab silent
21119 ORF1ab C2551F
21137 ORF1ab K2557R
21307 ORF1ab E2614stop
21403 ORF1ab del 1
22716 S T385I
23271 S A570D
23403 S D614G
26735 M silent
27033 M A171S
27972 ORF8 Q27stop
28048 ORF8 R52I
28093 ORF8 S67F
28111 ORF8 Y73C
28280 N D3L
28281 N D3L
28282 N D3L
2235
1596
1004
1778
187
1146
1341
923
931
1284
6468
1167
2077
2175
2256
2100
3697
3925
8468
8378
21866
12927
13610
10355
741
741
757
VSP0700−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

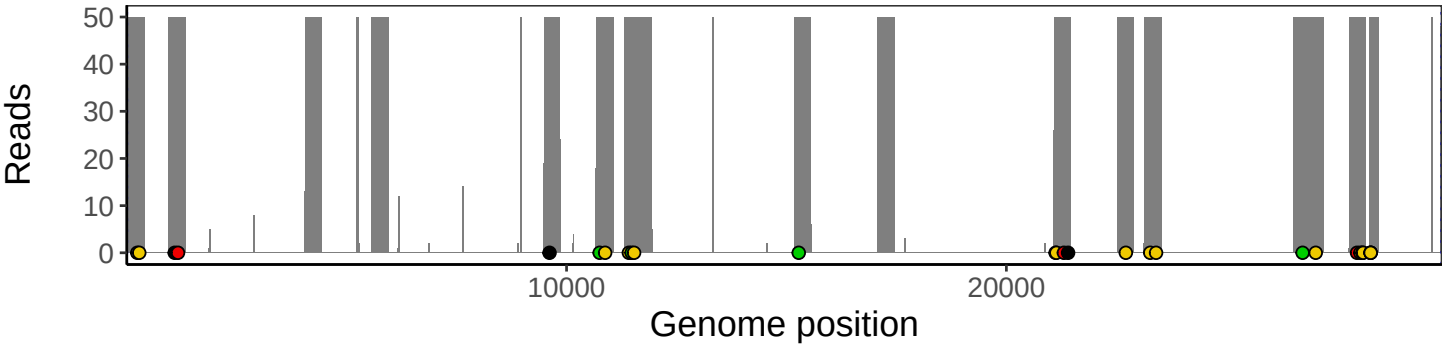
Analyses of individual experiments and composite results

VSP0700-1 | 2021-03-17 | Saline | H2103080875 | genomes | single experiment

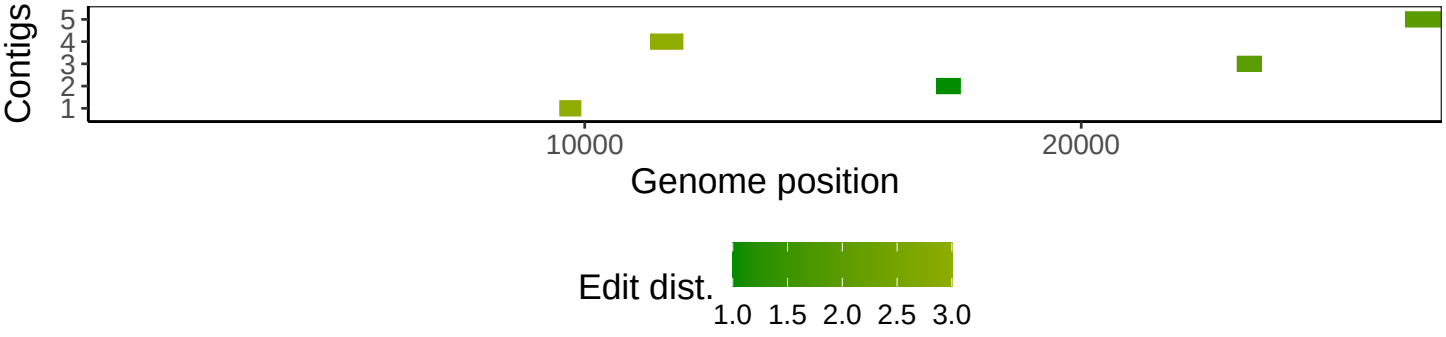
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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