COVID-19 subject UPHS-1522

2021-06-23

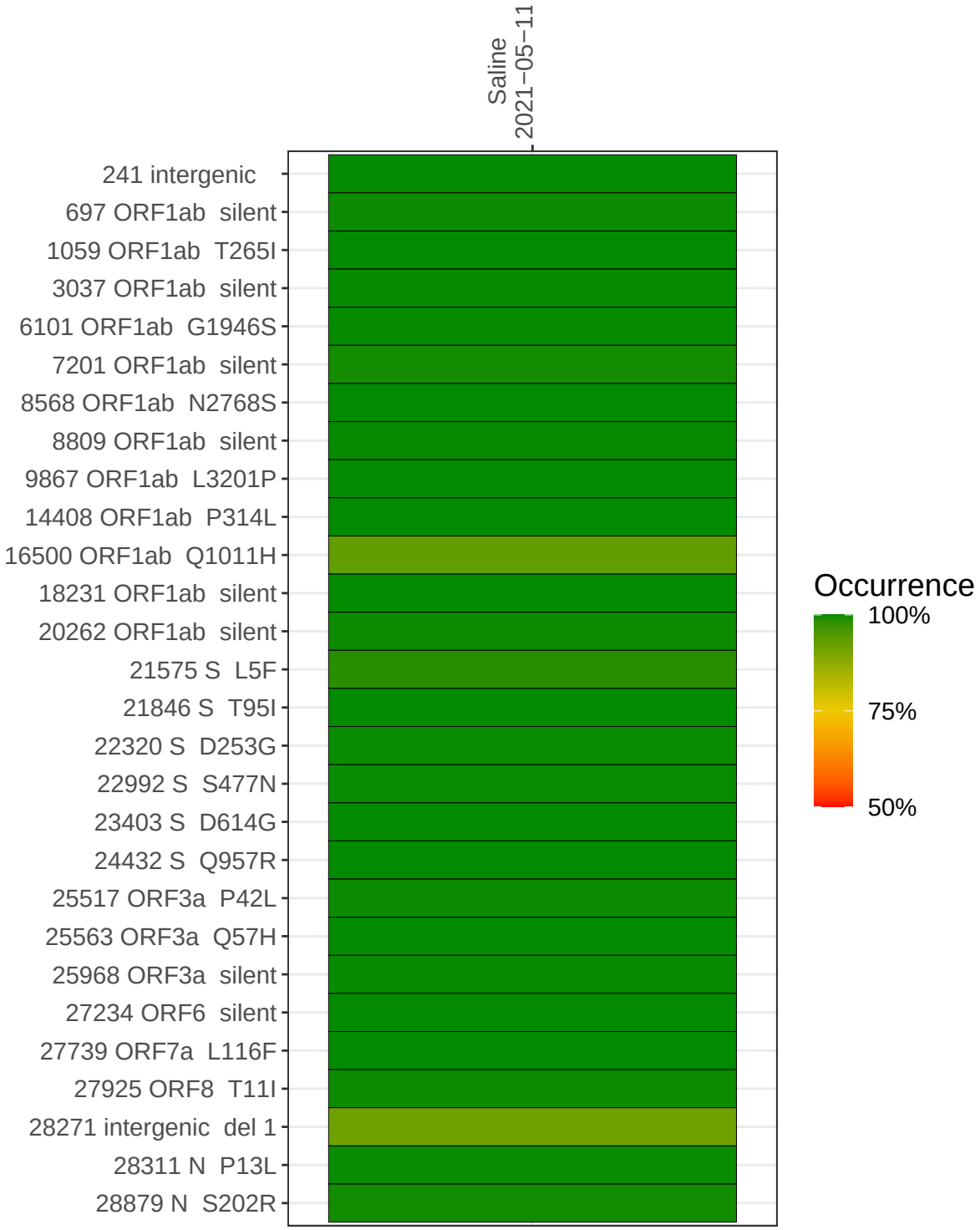
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

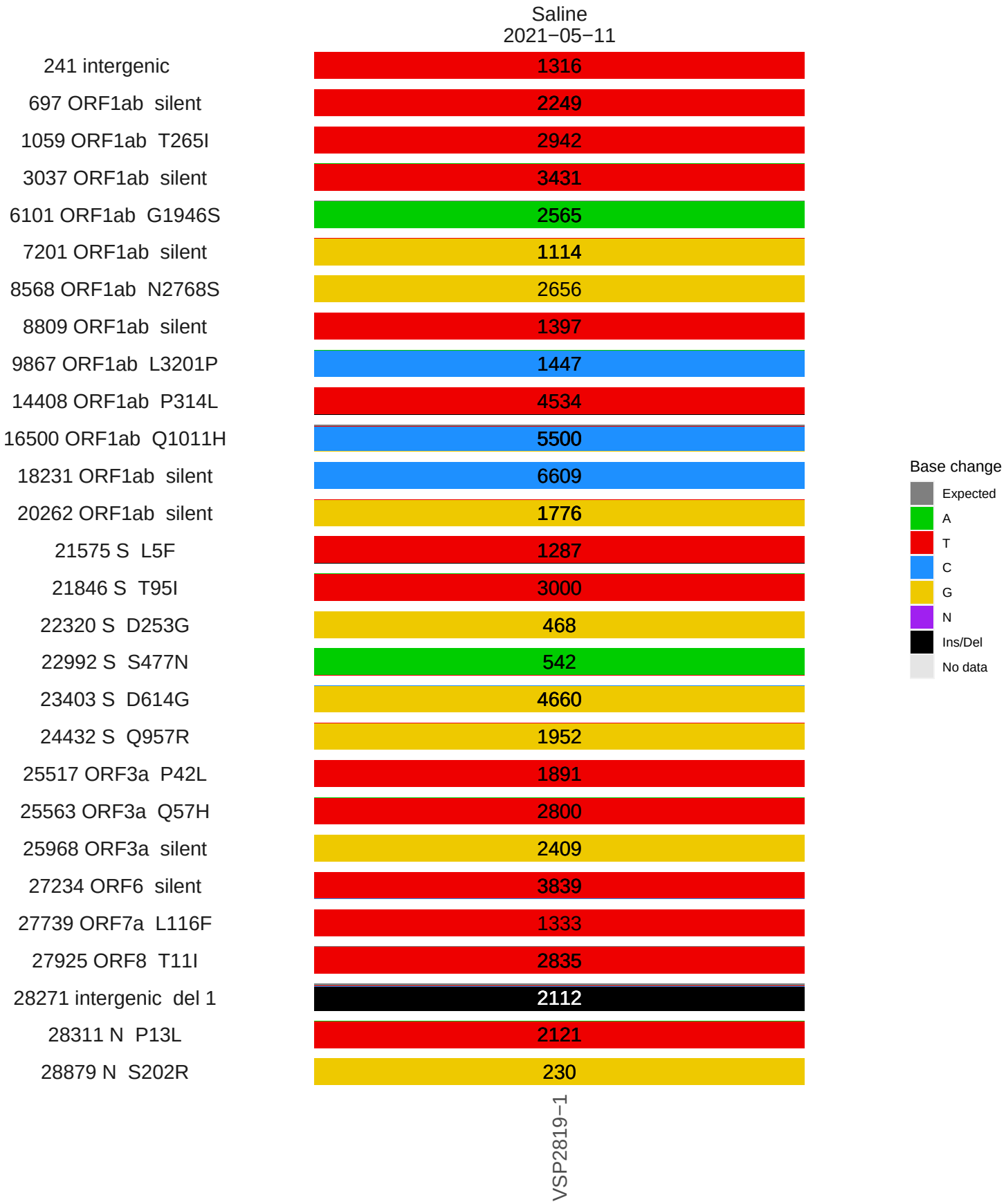
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP2819-1	single experiment	NA	Saline	2021-05-11	29.79	B.1	99.7%	99.7%

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.

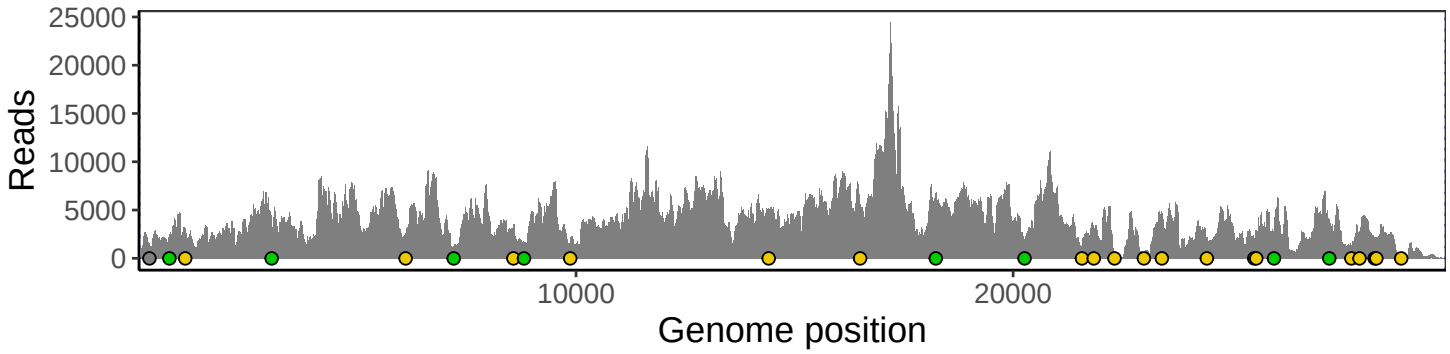



Saline
2021−05−11
241 intergenic
1316
697 ORF1ab silent
2249
1059 ORF1ab T265I
2942
3037 ORF1ab silent
3431
6101 ORF1ab G1946S
2565
7201 ORF1ab silent
1114
8568 ORF1ab N2768S
2656
8809 ORF1ab silent
1397
9867 ORF1ab L3201P
1447
14408 ORF1ab P314L
4534
16500 ORF1ab Q1011H
5500
18231 ORF1ab silent
6609
20262 ORF1ab silent
1776
21575 S L5F
1287
21846 S T95I
3000
22320 S D253G
468
22992 S S477N
542
23403 S D614G
4660
24432 S Q957R
1952
25517 ORF3a P42L
1891
25563 ORF3a Q57H
2800
25968 ORF3a silent
2409
27234 ORF6 silent
3839
27739 ORF7a L116F
1333
27925 ORF8 T11I
2835
28271 intergenic del 1
2112
28311 N P13L
2121
28879 N S202R
230
VSP2819−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

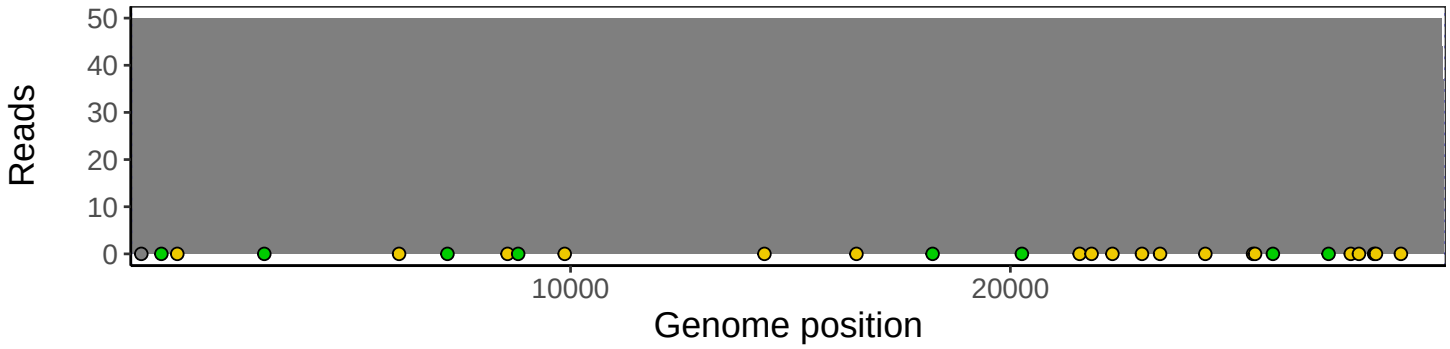
Analyses of individual experiments and composite results

VSP2819-1 | 2021-05-11 | Saline | UPHS-1522 | genomes | single experiment

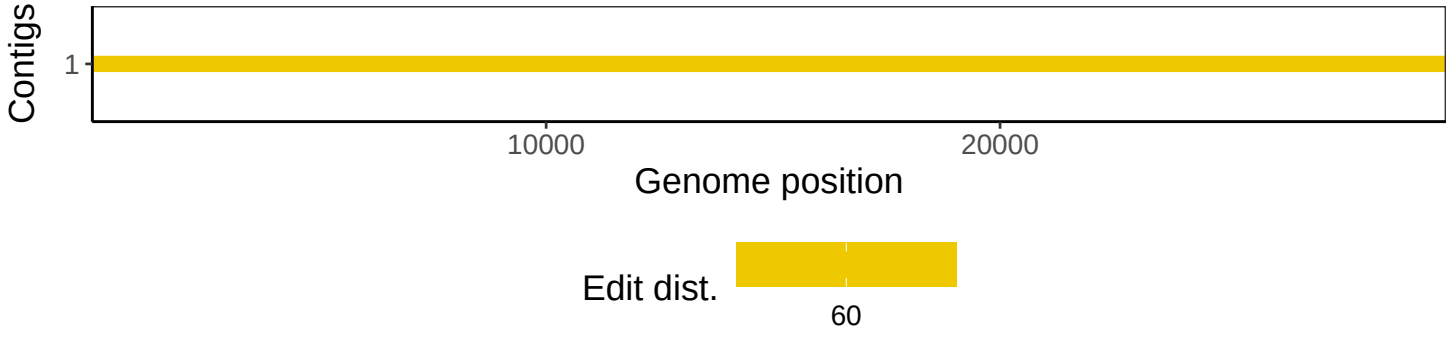
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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