COVID-19 subject UPHS-0685

2021-06-23

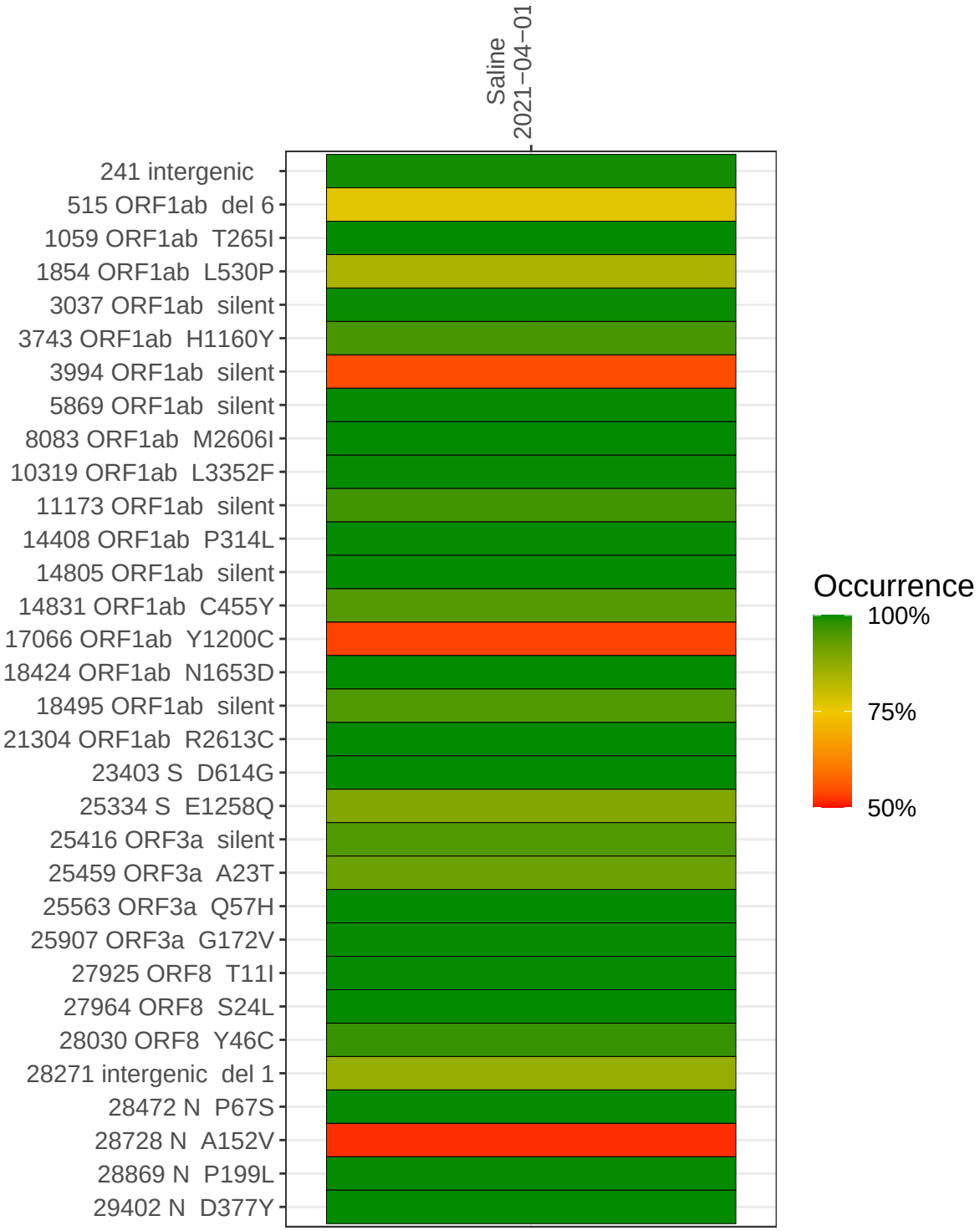
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

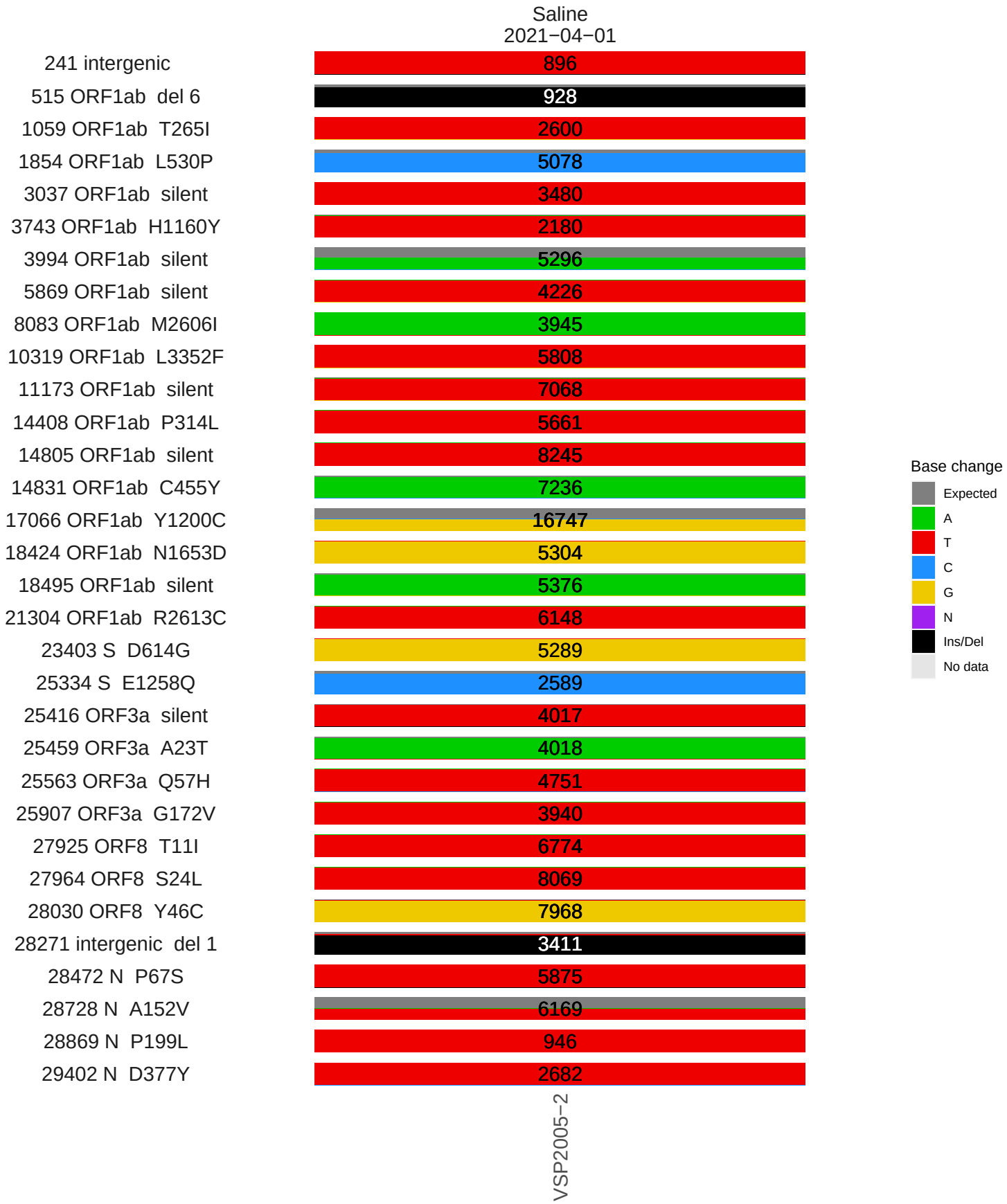
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP2005-2	single experiment	NA	Saline	2021-04-01	29.83	B.1.2	99.8%	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.

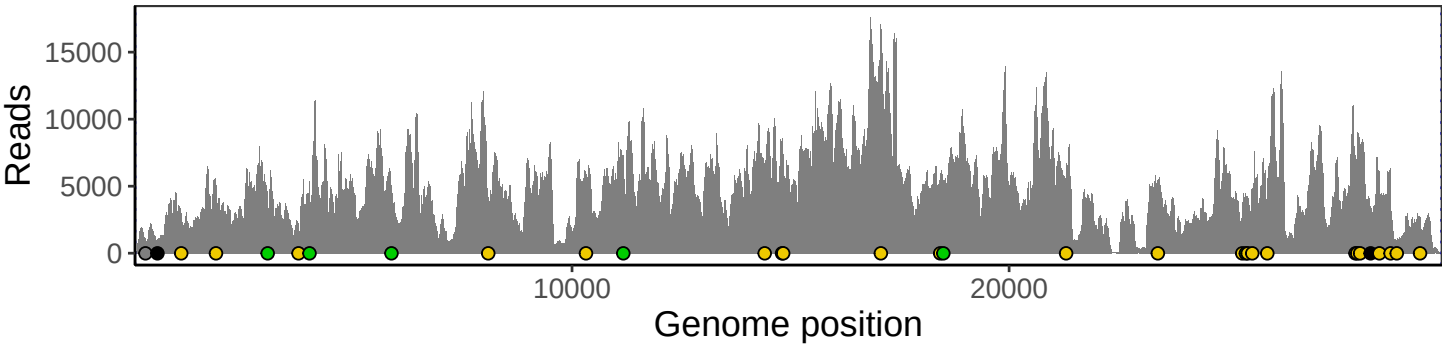



Saline
2021−04−01
241 intergenic
515 ORF1ab del 6
1059 ORF1ab T265I
1854 ORF1ab L530P
3037 ORF1ab silent
3743 ORF1ab H1160Y
3994 ORF1ab silent
5869 ORF1ab silent
8083 ORF1ab M2606I
10319 ORF1ab L3352F
11173 ORF1ab silent
14408 ORF1ab P314L
14805 ORF1ab silent
14831 ORF1ab C455Y
17066 ORF1ab Y1200C
18424 ORF1ab N1653D
18495 ORF1ab silent
21304 ORF1ab R2613C
23403 S D614G
25334 S E1258Q
25416 ORF3a silent
25459 ORF3a A23T
25563 ORF3a Q57H
25907 ORF3a G172V
27925 ORF8 T11I
27964 ORF8 S24L
28030 ORF8 Y46C
28271 intergenic del 1
28472 N P67S
28728 N A152V
28869 N P199L
29402 N D377Y
896
928
2600
5078
3480
2180
5296
4226
3945
5808
7068
5661
8245
7236
16747
5304
5376
6148
5289
2589
4017
4018
4751
3940
6774
8069
7968
3411
5875
6169
946
2682
VSP2005−2
Base change
Expected
A
T
C
G
N
Ins/Del
No data

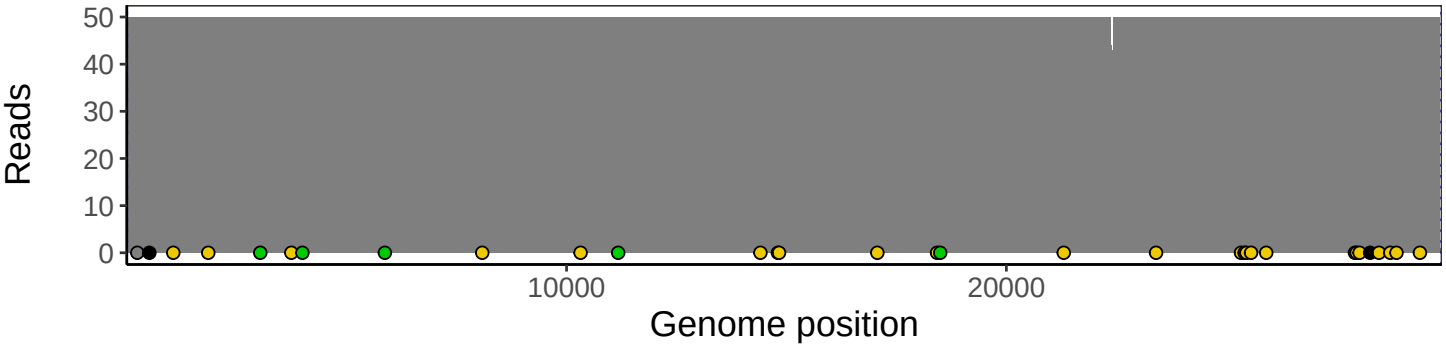
Analyses of individual experiments and composite results

VSP2005-2 | 2021-04-01 | Saline | UPHS-0685 | genomes | single experiment

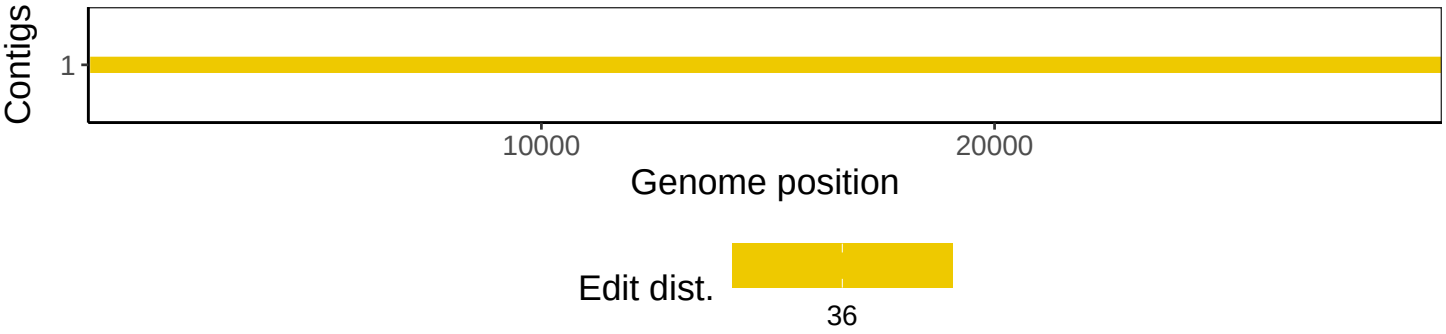
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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