COVID-19 subject UPHS-0648

2021-06-03

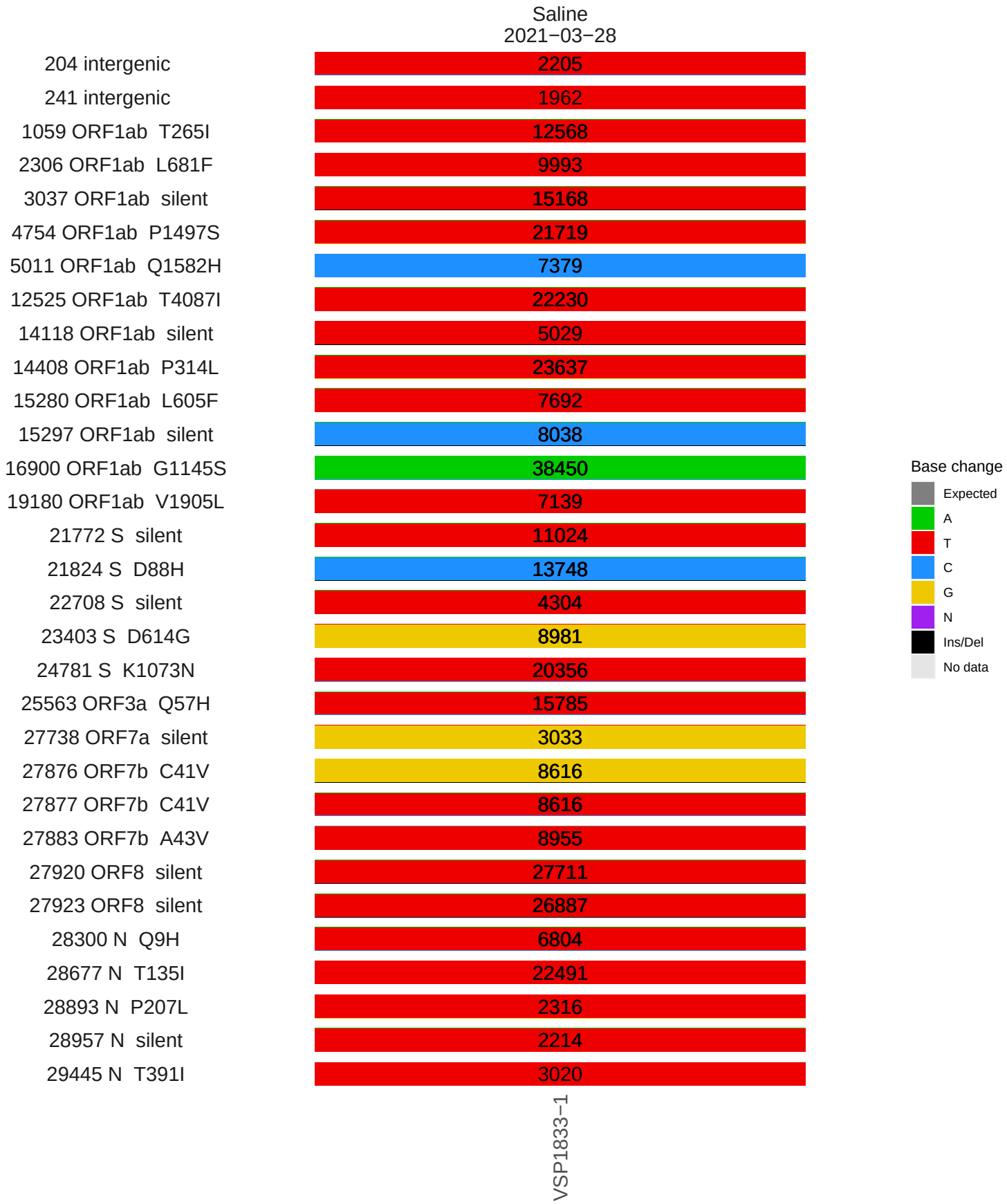
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP1833-1	single experiment	NA	Saline	2021-03-28	29.87	B.1.311	99.9%	99.8%

Variants shared across samples

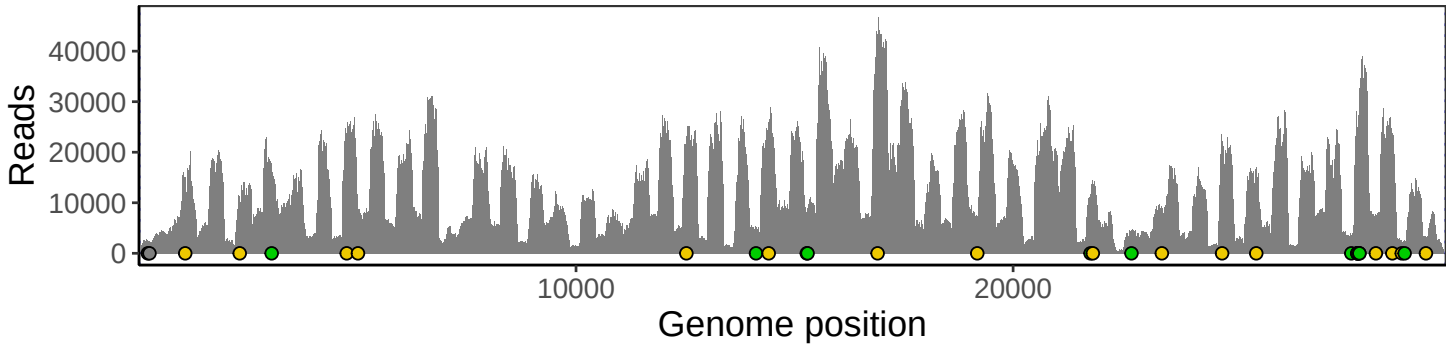
The heat map below shows how variants (reference genome /home/common/SARS-CoV-2-Philadelphia/Wuhan-Hu-1) are shared across subject samples where the percent variance is colored. Variants are called if a variant position is covered by 5 or more reads, the alternative base is found in $> 50\%$ of read 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
2021−03−28
204 intergenic
241 intergenic
1059 ORF1ab T265I
2306 ORF1ab L681F
3037 ORF1ab silent
4754 ORF1ab P1497S
5011 ORF1ab Q1582H
12525 ORF1ab T4087I
14118 ORF1ab silent
14408 ORF1ab P314L
15280 ORF1ab L605F
15297 ORF1ab silent
16900 ORF1ab G1145S
19180 ORF1ab V1905L
21772 S silent
21824 S D88H
22708 S silent
23403 S D614G
24781 S K1073N
25563 ORF3a Q57H
27738 ORF7a silent
27876 ORF7b C41V
27877 ORF7b C41V
27883 ORF7b A43V
27920 ORF8 silent
27923 ORF8 silent
28300 N Q9H
28677 N T135I
28893 N P207L
28957 N silent
29445 N T391I
2205
1962
12568
9993
15168
21719
7379
22230
5029
23637
7692
8038
38450
7139
11024
13748
4304
8981
20356
15785
3033
8616
8616
8955
27711
26887
6804
22491
2316
2214
3020
VSP1833−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

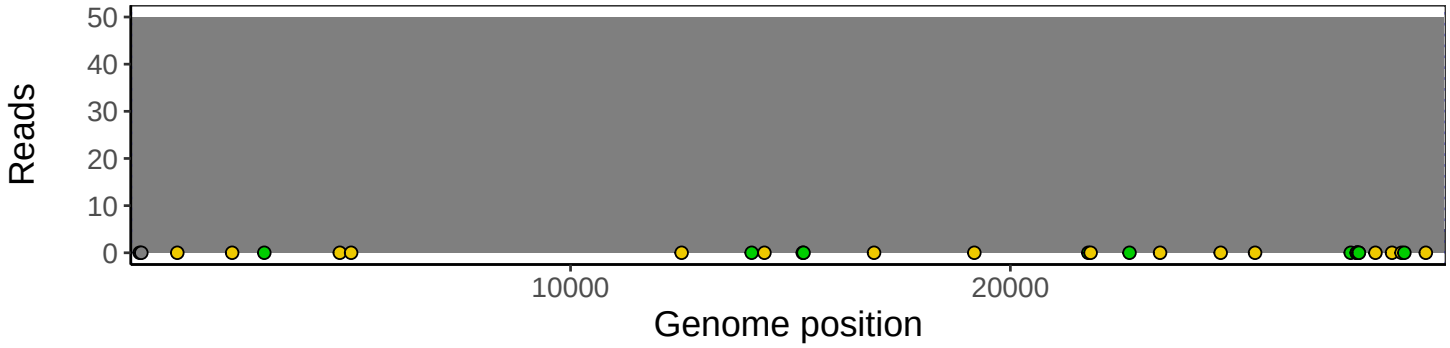
Analyses of individual experiments and composite results

VSP1833-1 | 2021-03-28 | Saline | UPHS-0648 | genomes | single experiment

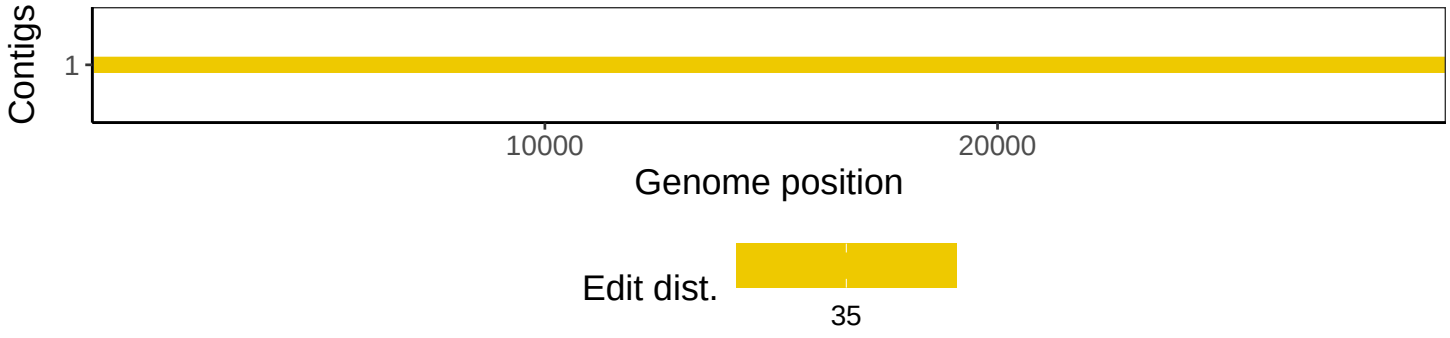
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	3.3.3
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