COVID-19 subject UPHS-0031

2021-04-17

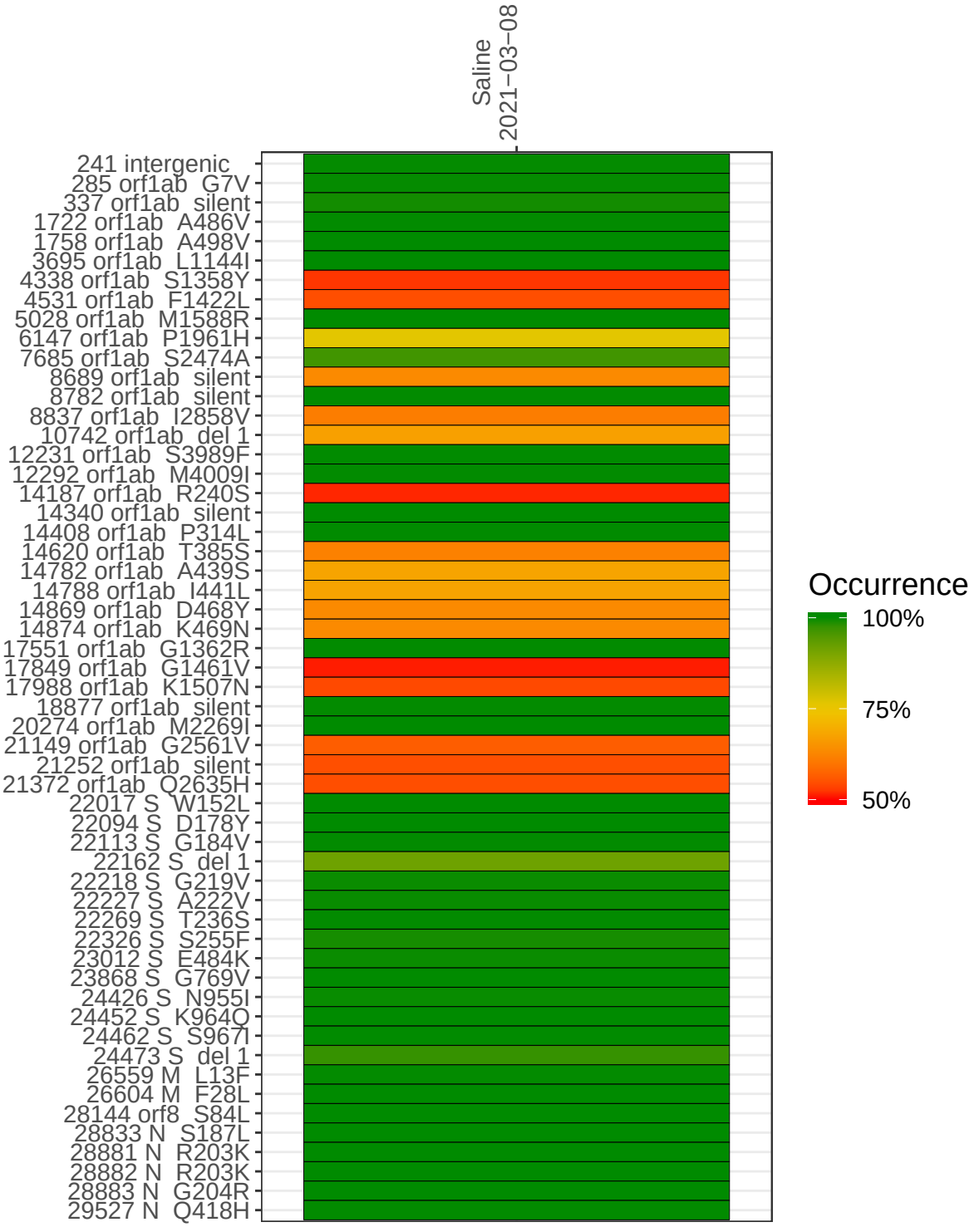
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

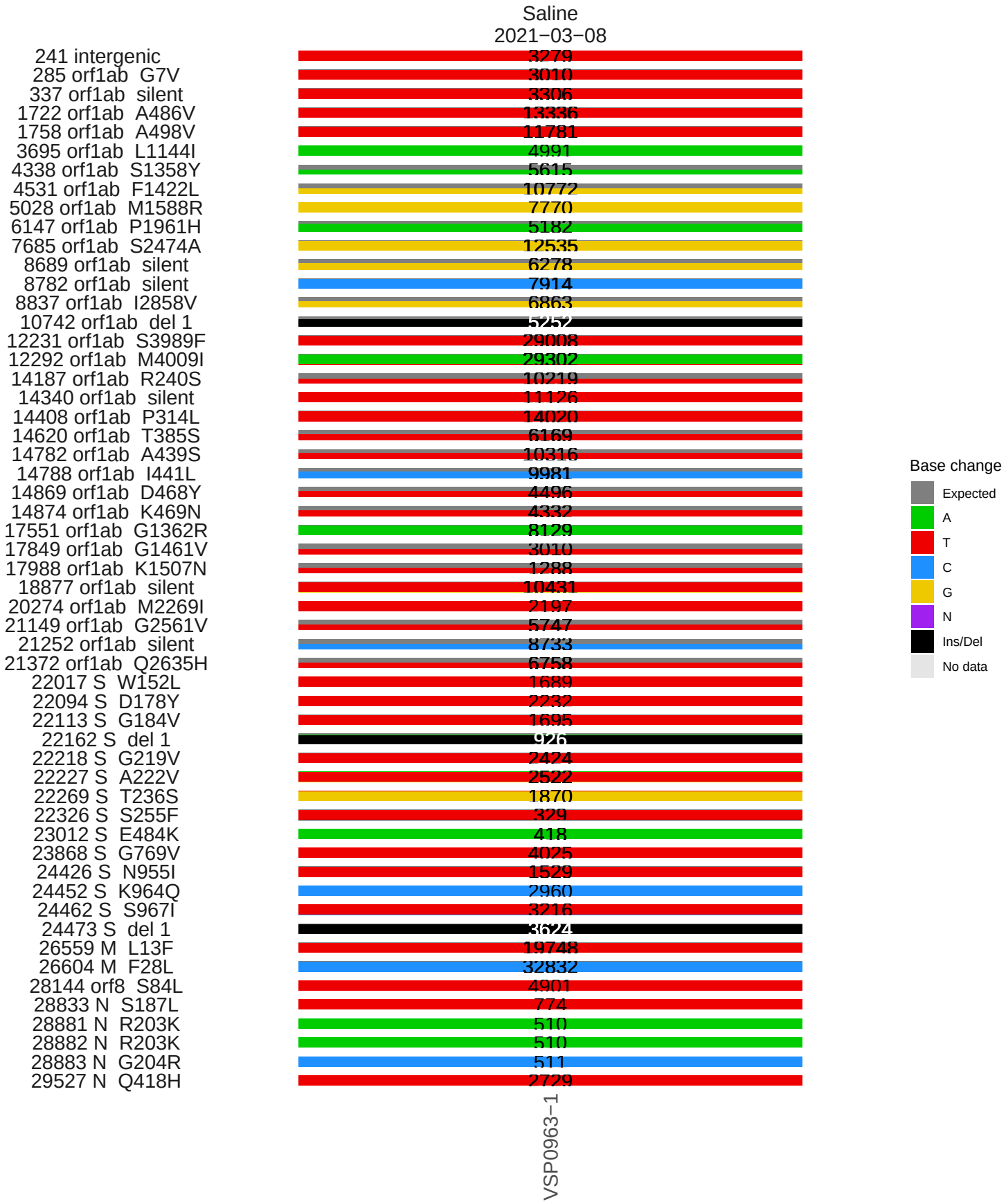
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0963-1	single experiment	NA	Saline	2021-03-08	6.45	NA	74.1%	73.7%

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.

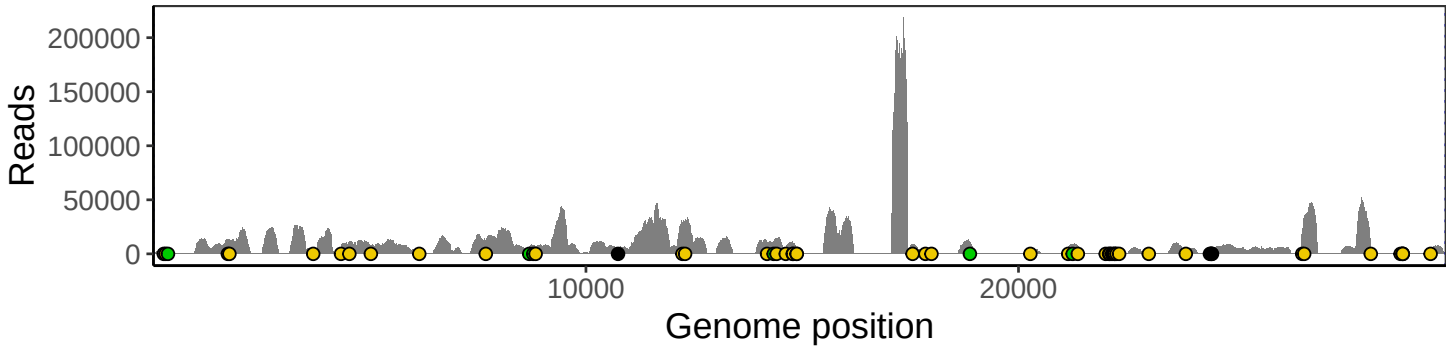



Saline
2021−03−08
241 intergenic 3279
285 orf1ab G7V 3010
337 orf1ab silent 3306
1722 orf1ab A486V 13336
1758 orf1ab A498V 11781
3695 orf1ab L1144I 4991
4338 orf1ab S1358Y 5615
4531 orf1ab F1422L 10772
5028 orf1ab M1588R 7770
6147 orf1ab P1961H 5182
7685 orf1ab S2474A 12535
8689 orf1ab silent 6278
8782 orf1ab silent 7914
8837 orf1ab I2858V 6863
10742 orf1ab del 1 5252
12231 orf1ab S3989F 29008
12292 orf1ab M4009I 29302
14187 orf1ab R240S 10219
14340 orf1ab silent 11126
14408 orf1ab P314L 14020
14620 orf1ab T385S 6169
14782 orf1ab A439S 10316
14788 orf1ab I441L 9981
14869 orf1ab D468Y 4496
14874 orf1ab K469N 4332
17551 orf1ab G1362R 8129
17849 orf1ab G1461V 3010
17988 orf1ab K1507N 1288
18877 orf1ab silent 10431
20274 orf1ab M2269I 2197
21149 orf1ab G2561V 5747
21252 orf1ab silent 8733
21372 orf1ab Q2635H 6758
22017 S W152L 1689
22094 S D178Y 2232
22113 S G184V 1695
22162 S del 1 926
22218 S G219V 2424
22227 S A222V 2522
22269 S T236S 1870
22326 S S255F 329
23012 S E484K 418
23868 S G769V 4025
24426 S N955I 1529
24452 S K964Q 2960
24462 S S967I 3216
24473 S del 1 3624
26559 M L13F 19748
26604 M F28L 32832
28144 orf8 S84L 4901
28833 N S187L 774
28881 N R203K 510
28882 N R203K 510
28883 N G204R 511
29527 N Q418H 2729
VSP0963−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

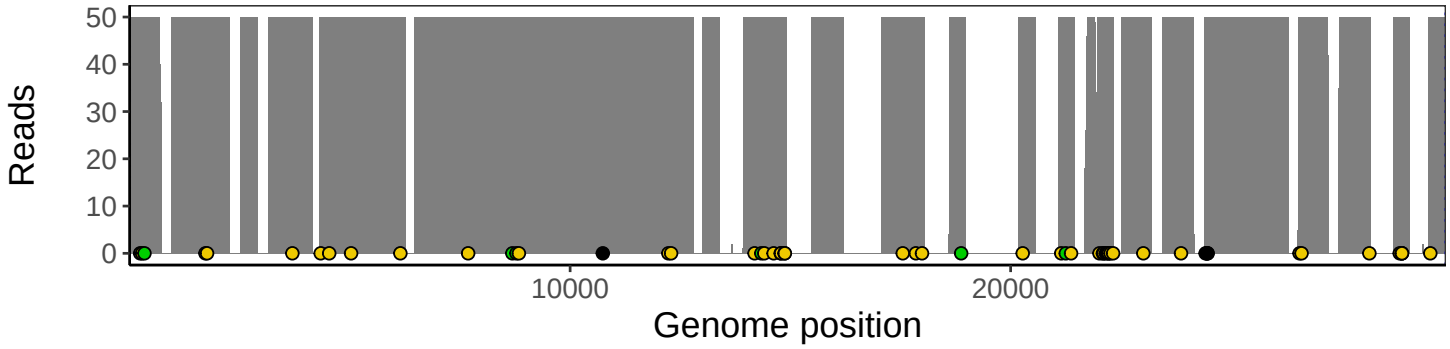
Analyses of individual experiments and composite results

VSP0963-1 | 2021-03-08 | Saline | UPHS-0031 | genomes | single experiment

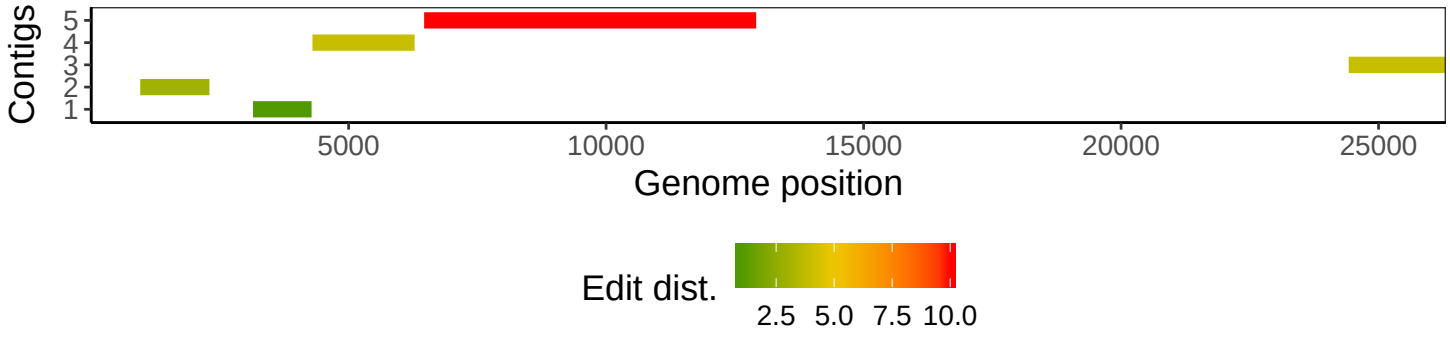
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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