COVID-19 subject UPHS-0102

2021-04-01

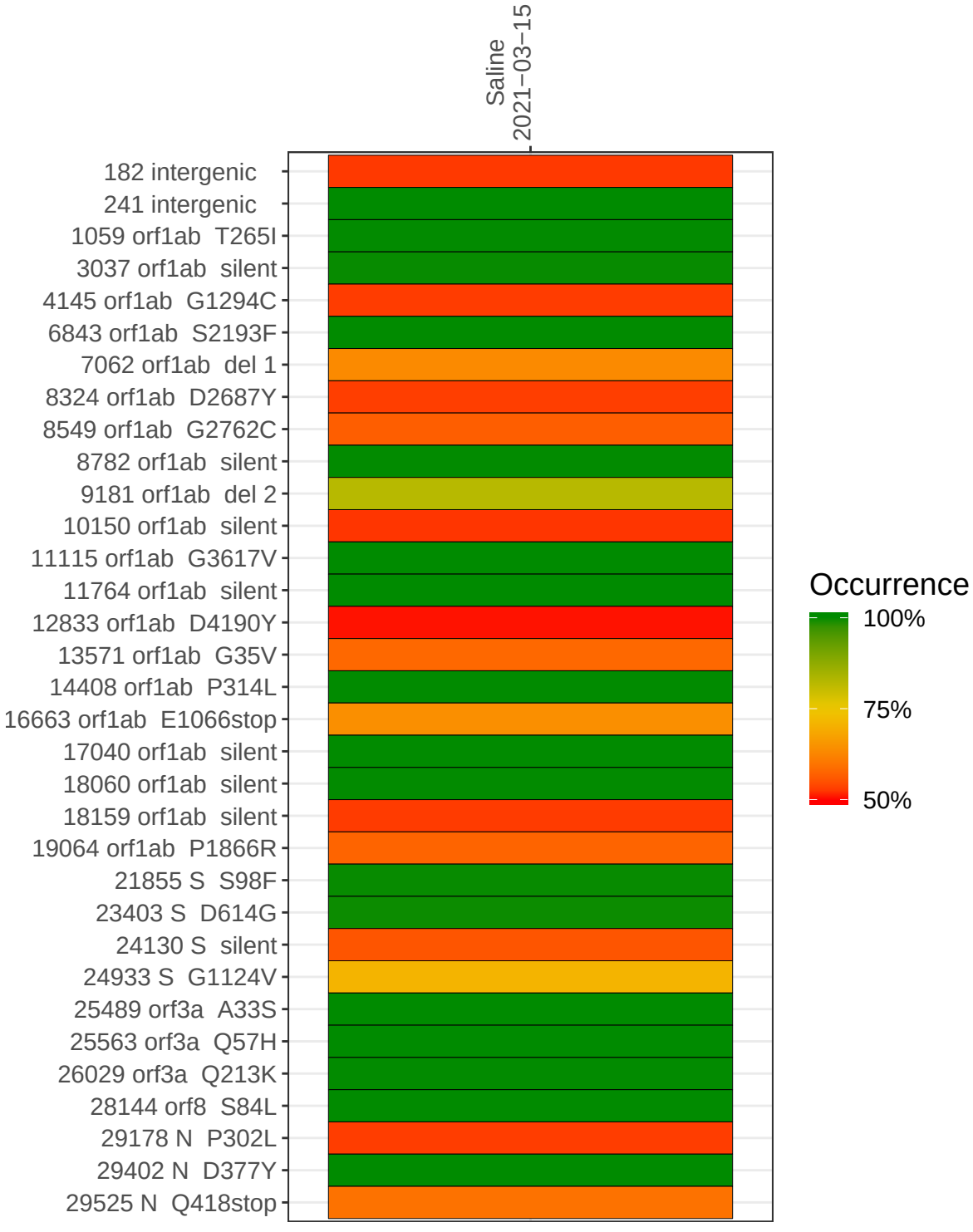
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

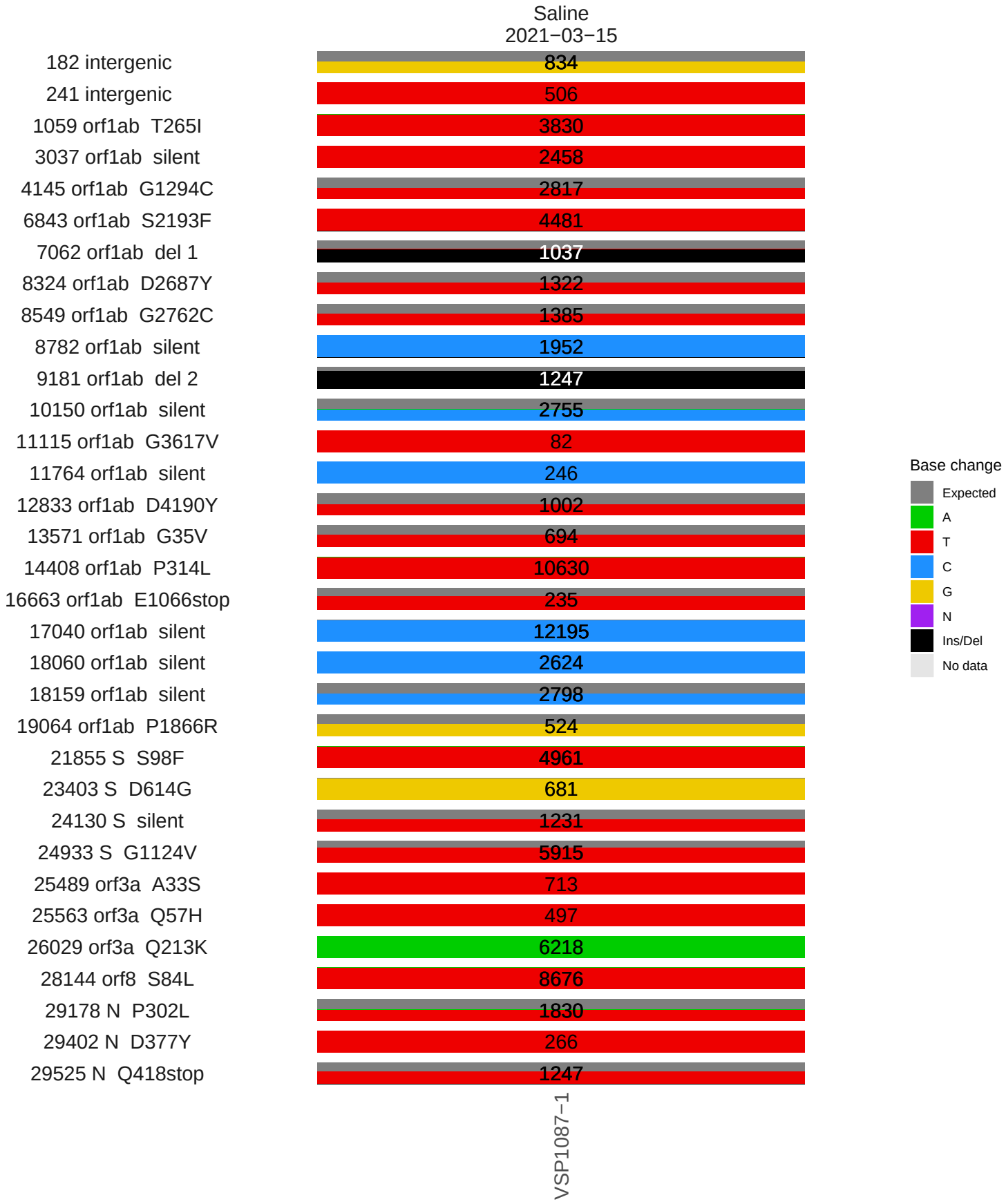
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1087-1	single experiment	NA	Saline	2021-03-15	11.30	B.1.433	95.3%	94.5%

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.

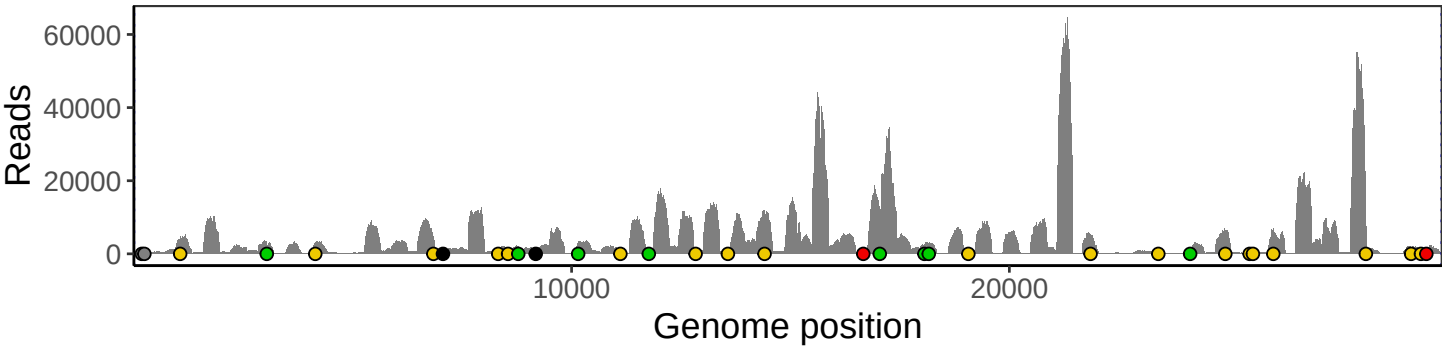



Saline
2021−03−15
182 intergenic
241 intergenic
1059 orf1ab T265I
3037 orf1ab silent
4145 orf1ab G1294C
6843 orf1ab S2193F
7062 orf1ab del 1
8324 orf1ab D2687Y
8549 orf1ab G2762C
8782 orf1ab silent
9181 orf1ab del 2
10150 orf1ab silent
11115 orf1ab G3617V
11764 orf1ab silent
12833 orf1ab D4190Y
13571 orf1ab G35V
14408 orf1ab P314L
16663 orf1ab E1066stop
17040 orf1ab silent
18060 orf1ab silent
18159 orf1ab silent
19064 orf1ab P1866R
21855 S S98F
23403 S D614G
24130 S silent
24933 S G1124V
25489 orf3a A33S
25563 orf3a Q57H
26029 orf3a Q213K
28144 orf8 S84L
29178 N P302L
29402 N D377Y
29525 N Q418stop
834
506
3830
2458
2817
4481
1037
1322
1385
1952
1247
2755
82
246
1002
694
10630
235
12195
2624
2798
524
4961
681
1231
5915
713
497
6218
8676
1830
266
1247
VSP1087−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

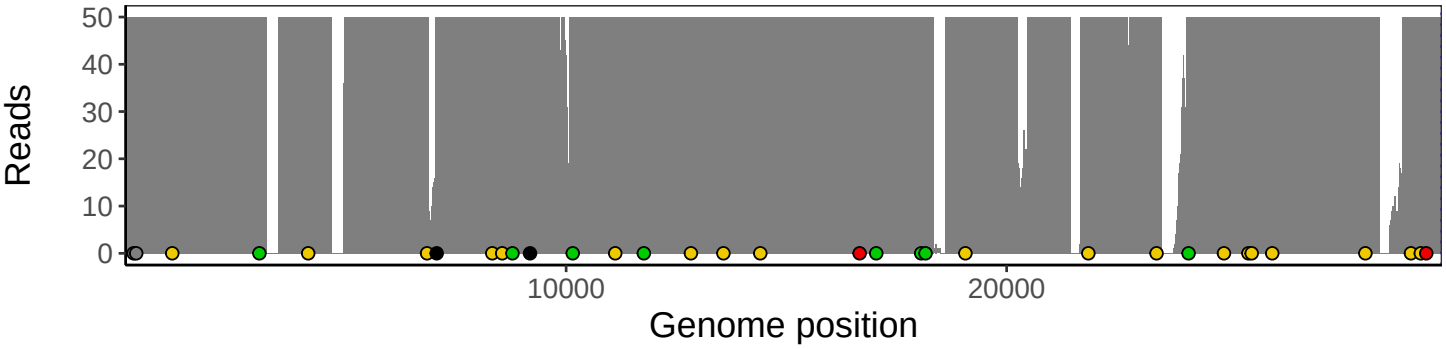
Analyses of individual experiments and composite results

VSP1087-1 | 2021-03-15 | Saline | UPHS-0102 | genomes | single experiment

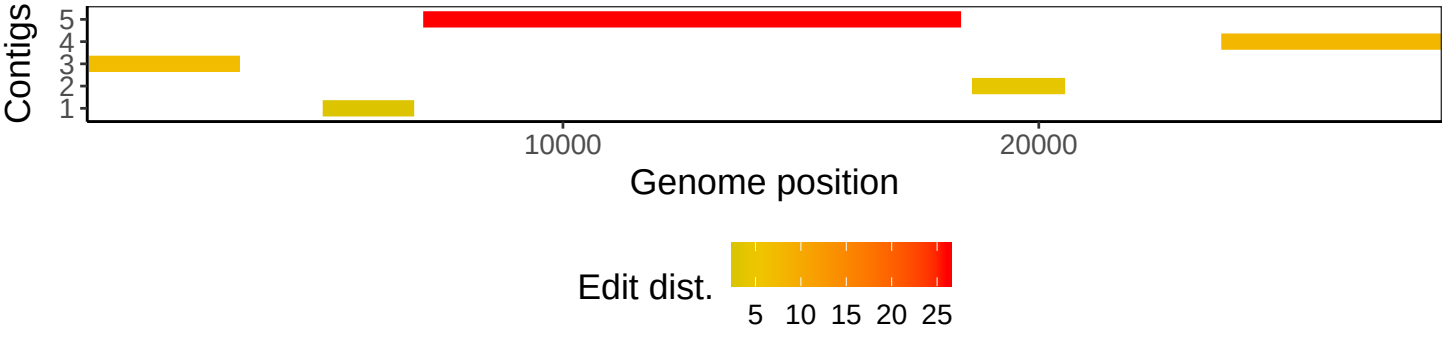
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.3.3
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	3.0.0
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