COVID-19 subject UPHS-0108

2021-03-29

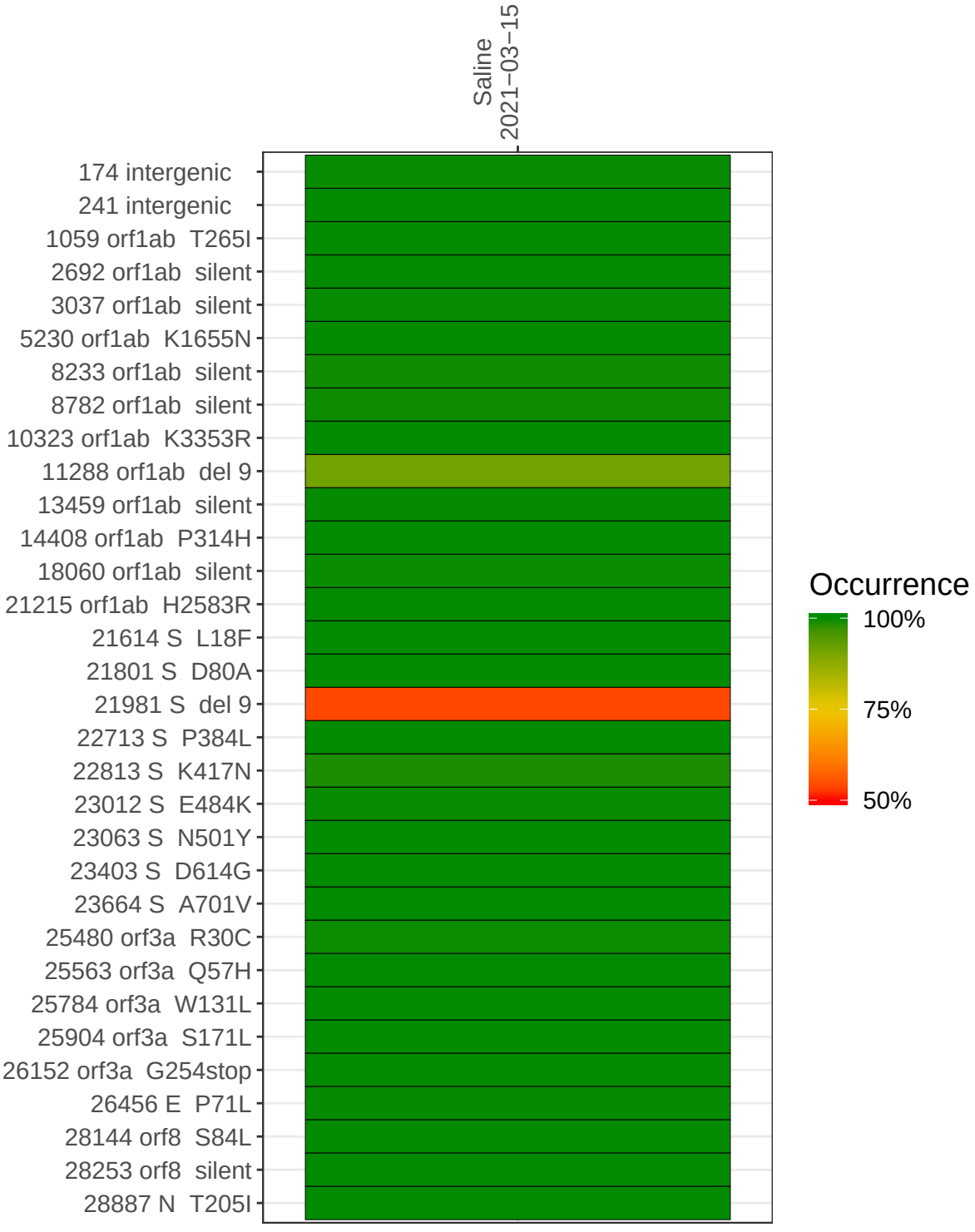
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

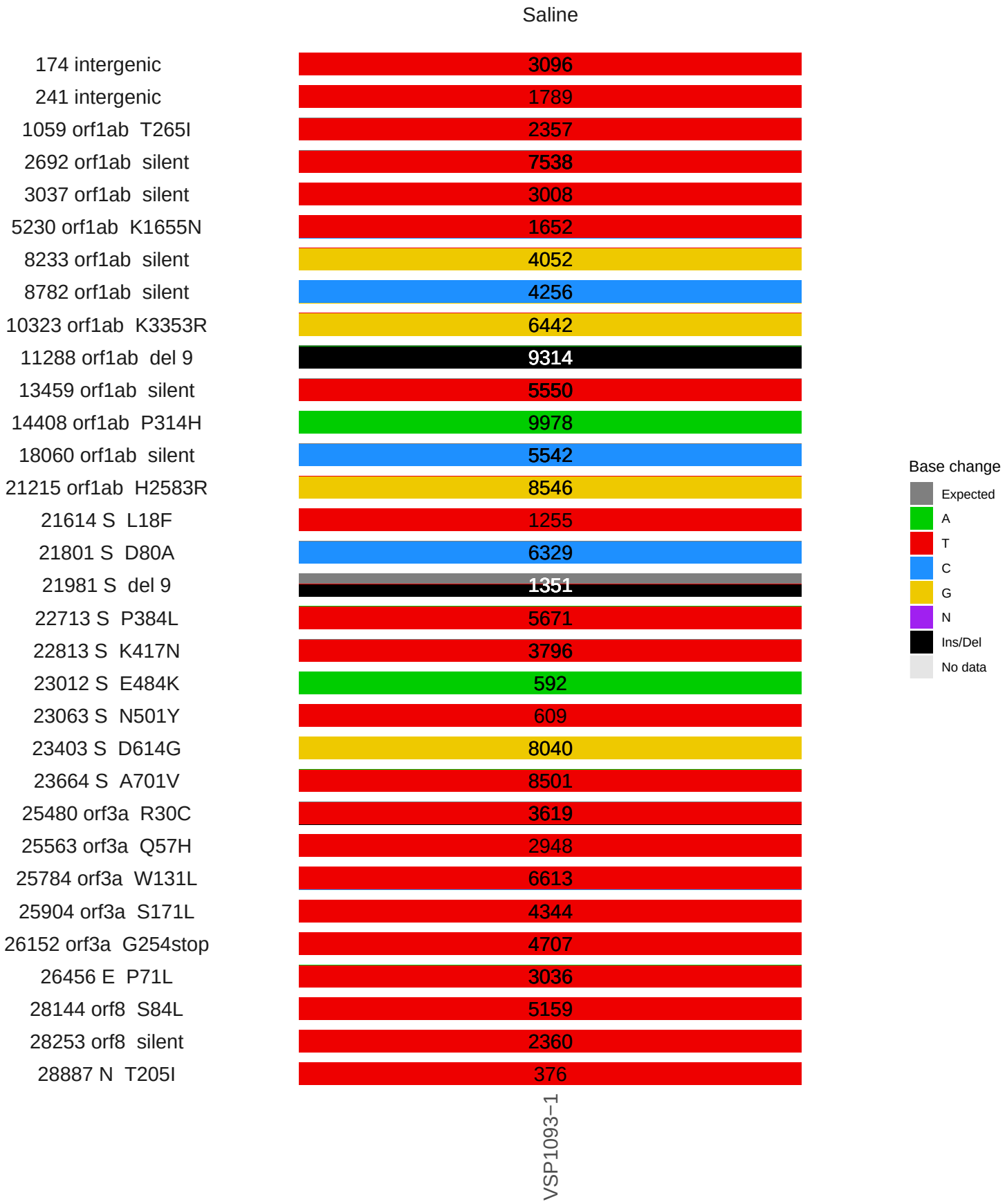
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1093-1	single experiment	NA	Saline	2021-03-15	21.97	B.1.351	99.1%	98.8%

Variants shared across samples

The heat map below shows how variants (reference genome 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.

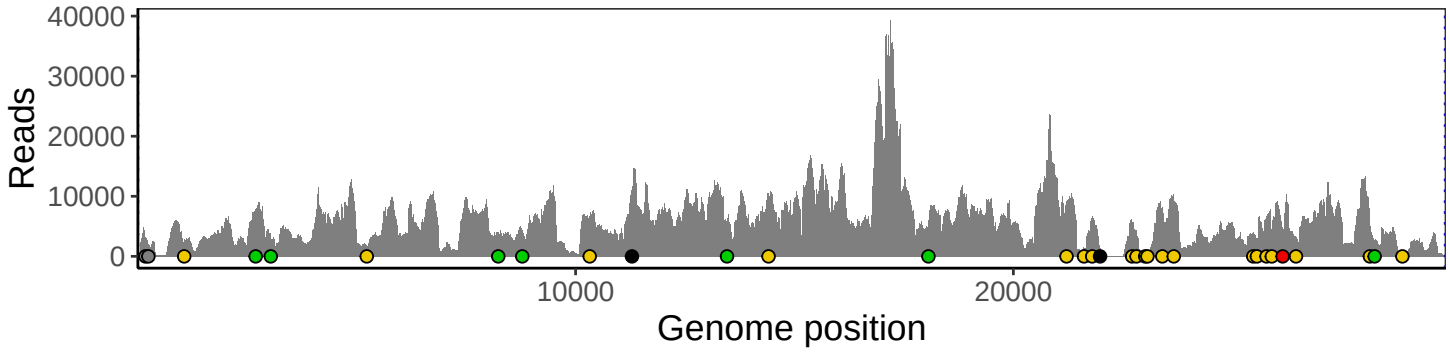



Saline
174 intergenic
241 intergenic
1059 orf1ab T265I
2692 orf1ab silent
3037 orf1ab silent
5230 orf1ab K1655N
8233 orf1ab silent
8782 orf1ab silent
10323 orf1ab K3353R
11288 orf1ab del 9
13459 orf1ab silent
14408 orf1ab P314H
18060 orf1ab silent
21215 orf1ab H2583R
21614 S L18F
21801 S D80A
21981 S del 9
22713 S P384L
22813 S K417N
23012 S E484K
23063 S N501Y
23403 S D614G
23664 S A701V
25480 orf3a R30C
25563 orf3a Q57H
25784 orf3a W131L
25904 orf3a S171L
26152 orf3a G254stop
26456 E P71L
28144 orf8 S84L
28253 orf8 silent
28887 N T205I
3096
1789
2357
7538
3008
1652
4052
4256
6442
9314
5550
9978
5542
8546
1255
6329
1351
5671
3796
592
609
8040
8501
3619
2948
6613
4344
4707
3036
5159
2360
376
VSP1093−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

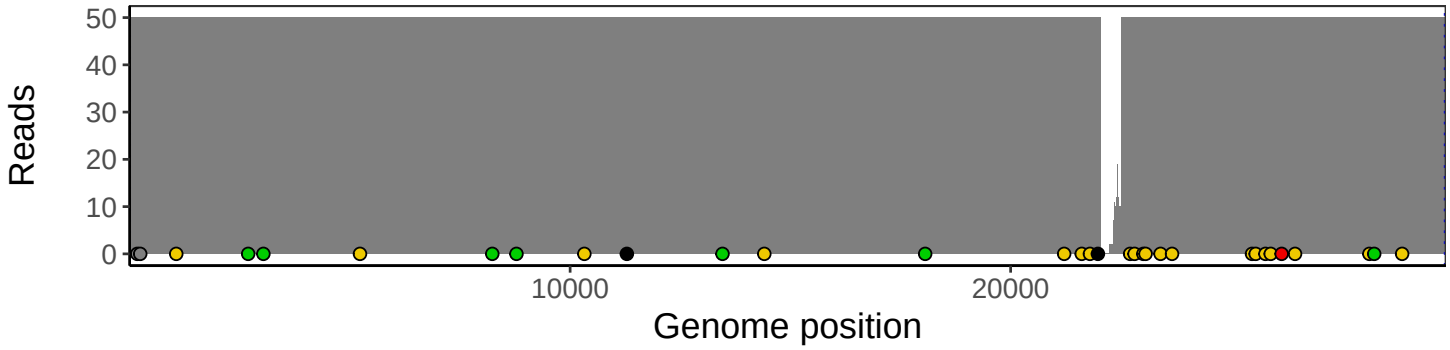
Analyses of individual experiments and composite results

VSP1093-1 | 2021-03-15 | Saline | UPHS-0108 | genomes | single experiment

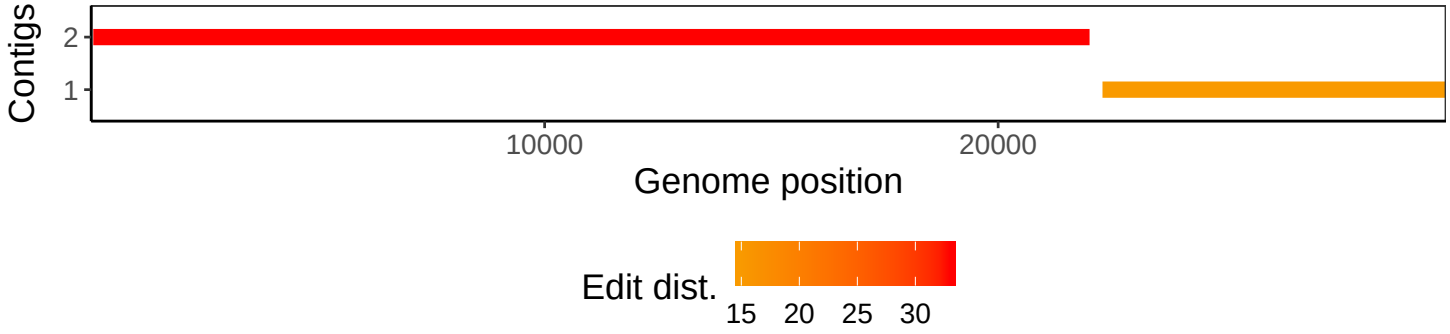
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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