COVID-19 subject UPHS-0090

2021-04-17

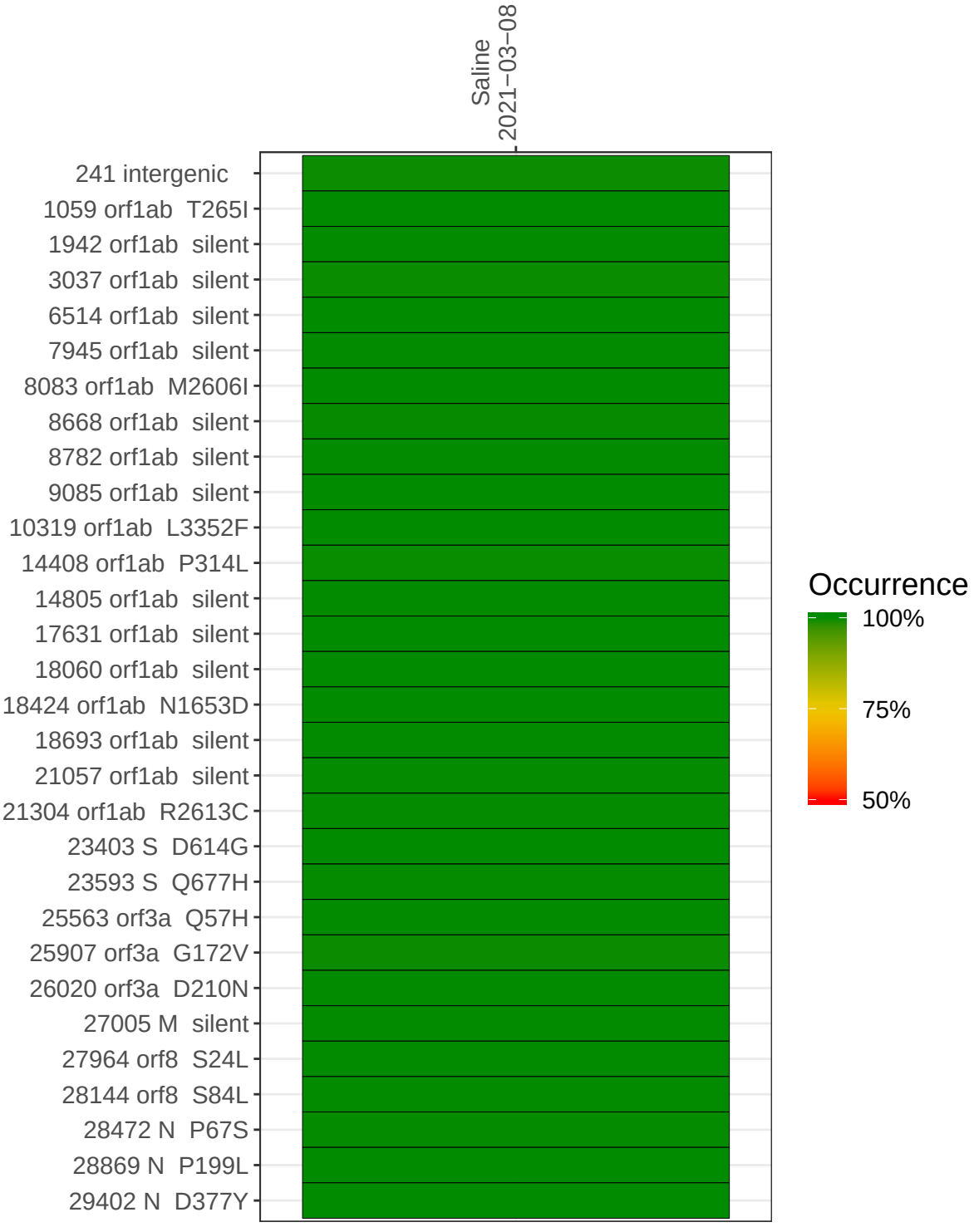
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

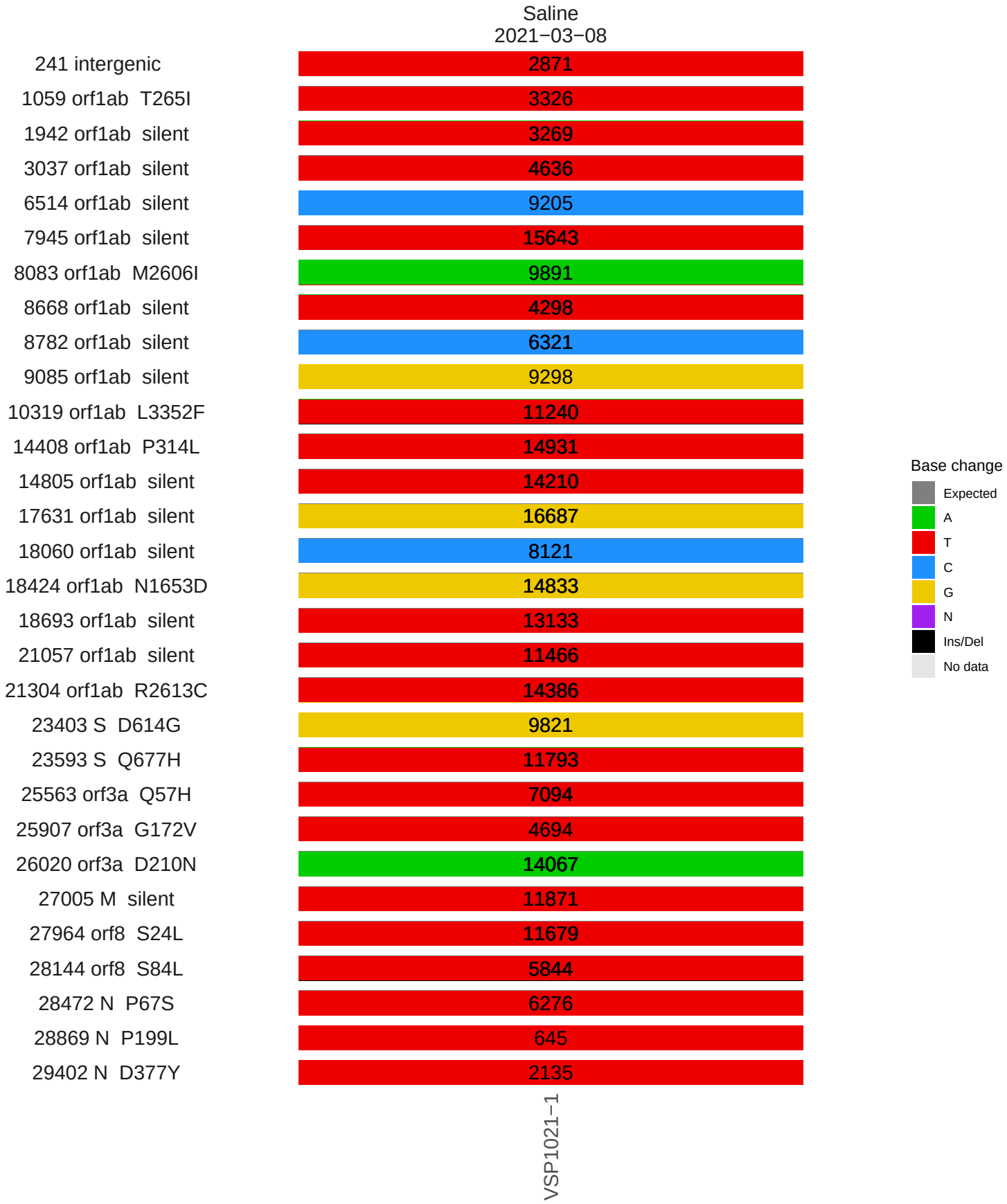
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1021-1	single experiment	NA	Saline	2021-03-08	29.93	B.1.2	99.9%	99.8%

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.

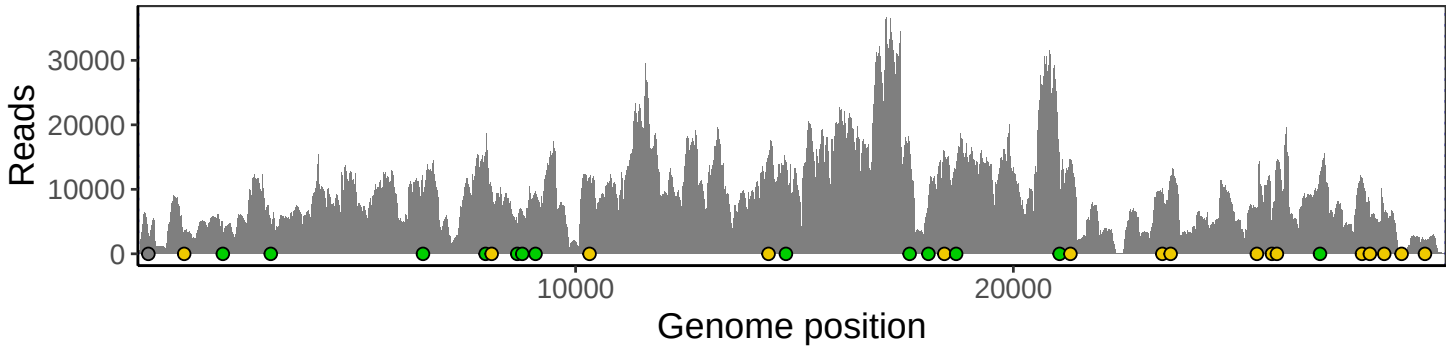



Saline
2021−03−08
241 intergenic
2871
1059 orf1ab T265I
3326
1942 orf1ab silent
3269
3037 orf1ab silent
4636
6514 orf1ab silent
9205
7945 orf1ab silent
15643
8083 orf1ab M2606I
9891
8668 orf1ab silent
4298
8782 orf1ab silent
6321
9085 orf1ab silent
9298
10319 orf1ab L3352F
11240
14408 orf1ab P314L
14931
14805 orf1ab silent
14210
17631 orf1ab silent
16687
18060 orf1ab silent
8121
18424 orf1ab N1653D
14833
18693 orf1ab silent
13133
21057 orf1ab silent
11466
21304 orf1ab R2613C
14386
23403 S D614G
9821
23593 S Q677H
11793
25563 orf3a Q57H
7094
25907 orf3a G172V
4694
26020 orf3a D210N
14067
27005 M silent
11871
27964 orf8 S24L
11679
28144 orf8 S84L
5844
28472 N P67S
6276
28869 N P199L
645
29402 N D377Y
2135
VSP1021−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

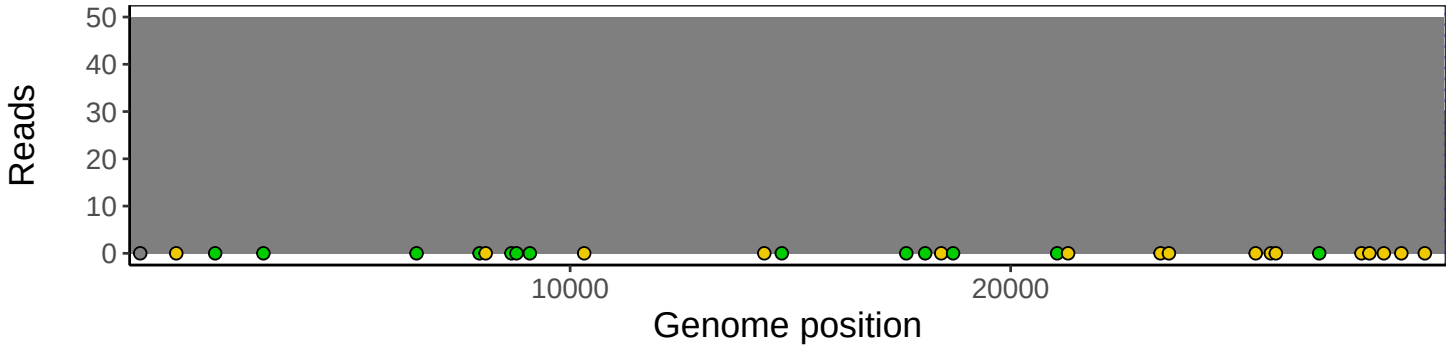
Analyses of individual experiments and composite results

VSP1021-1 | 2021-03-08 | Saline | UPHS-0090 | genomes | single experiment

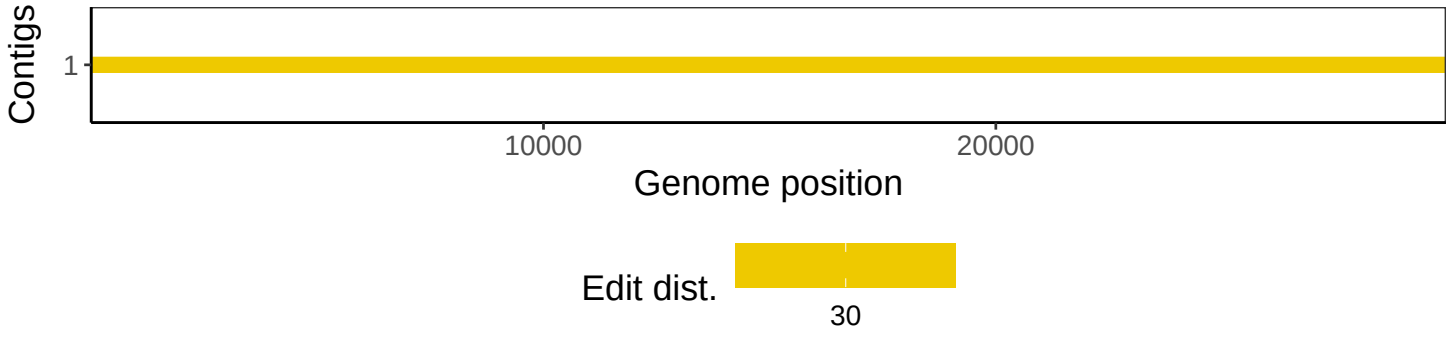
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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