COVID-19 subject UPHS-0104

2021-04-17

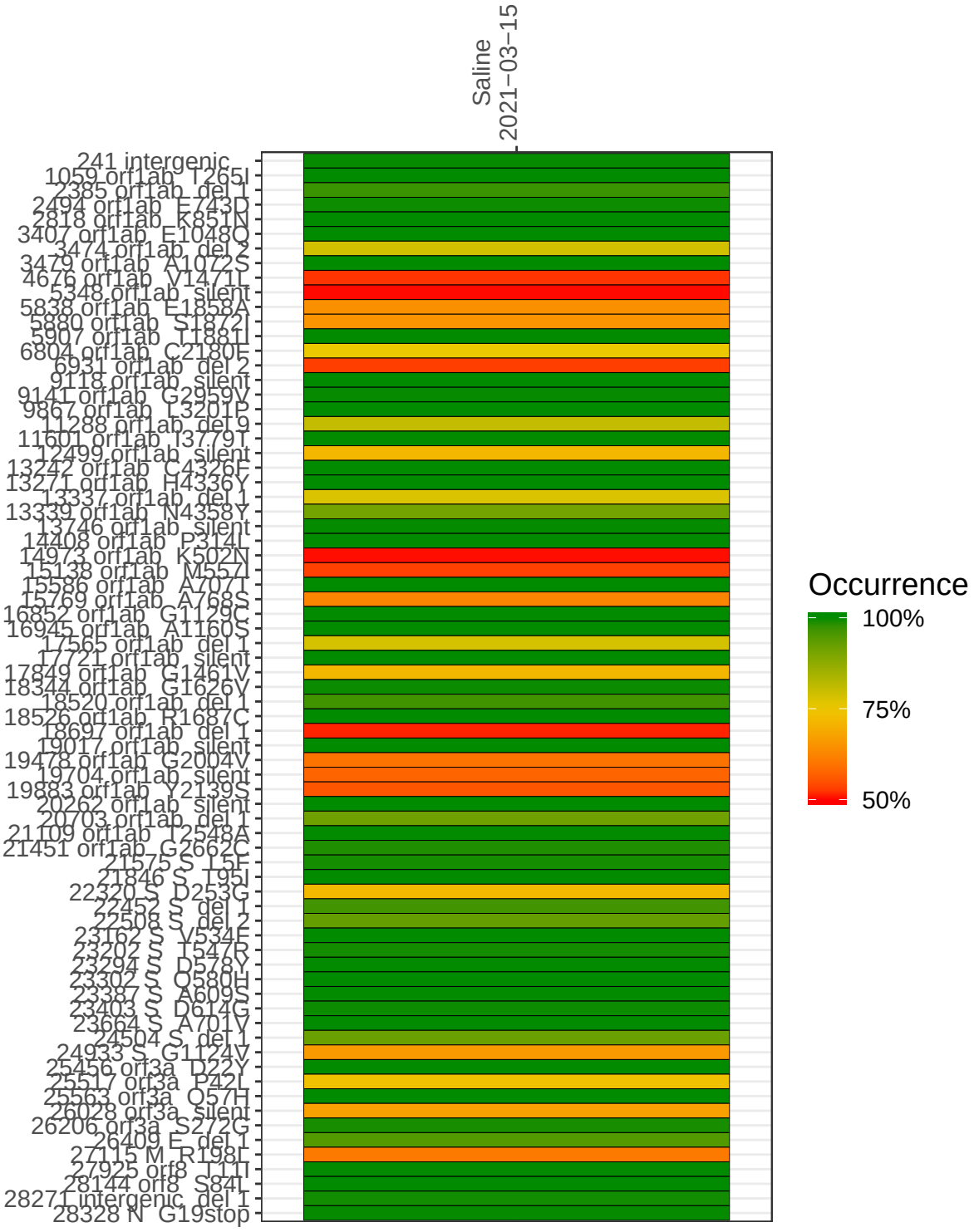
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

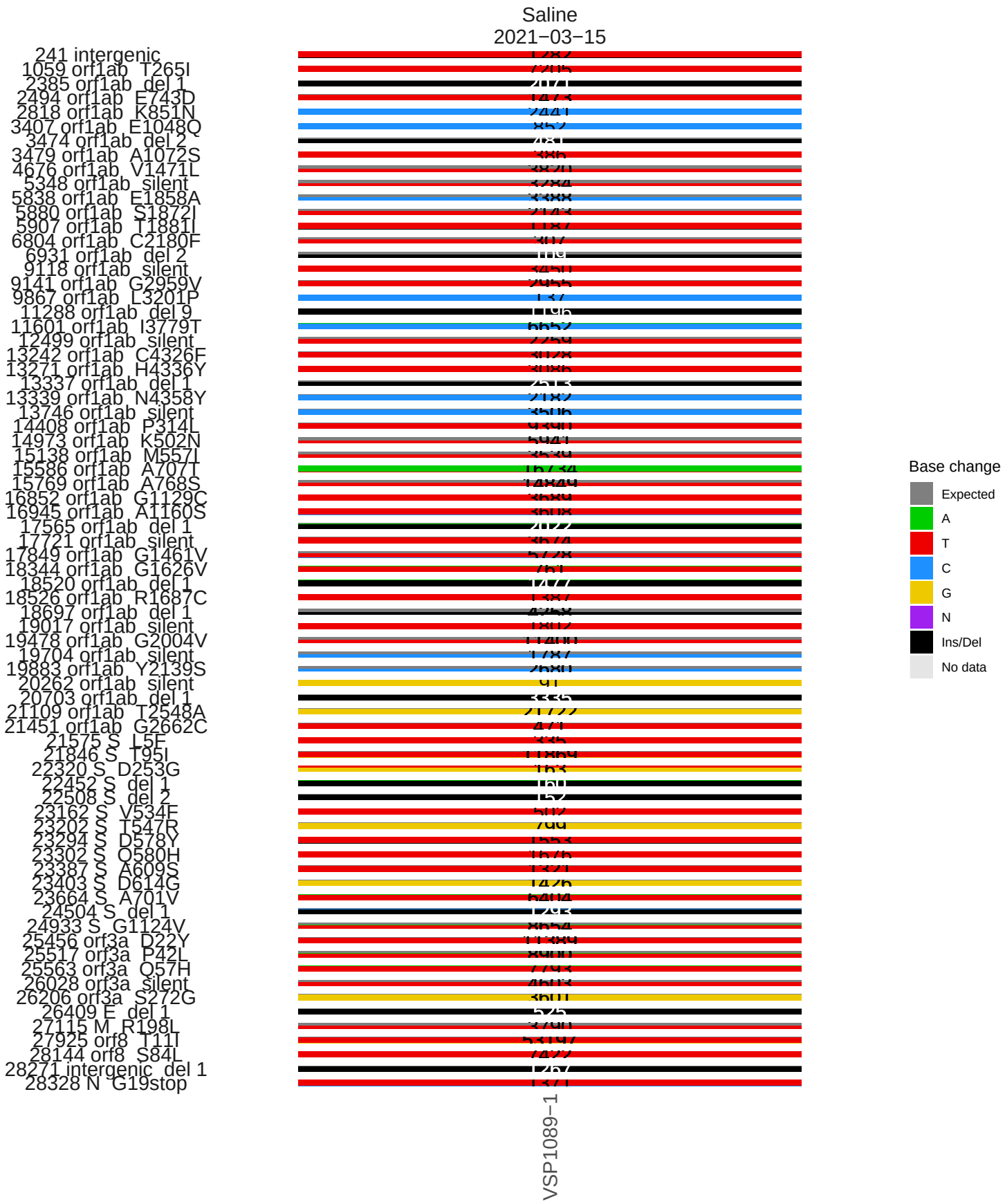
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1089-1	single experiment	NA	Saline	2021-03-15	4.63	NA	78.0%	76.9%

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.

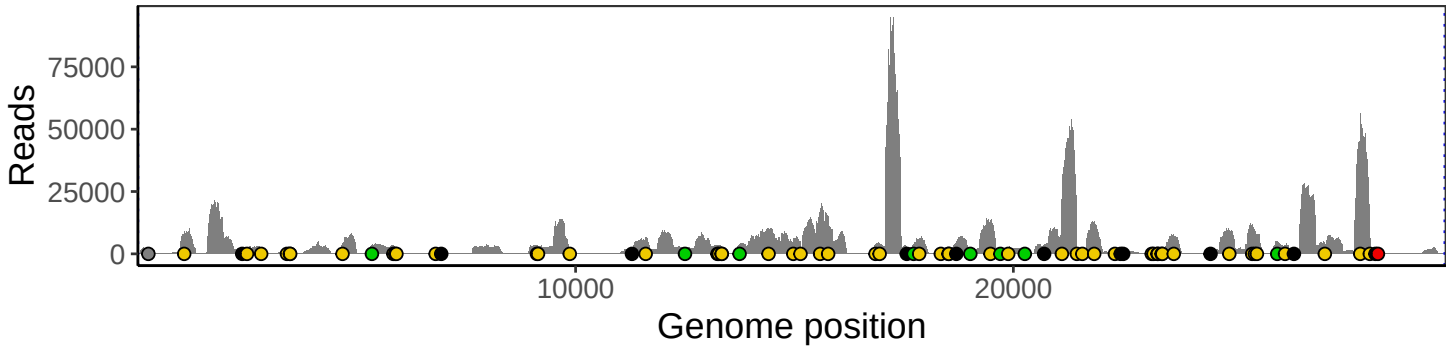



Saline
2021−03−15
241 intergenic
1059 orf1ab T265I
2385 orf1ab del 1
2494 orf1ab E743D
2818 orf1ab K851N
3407 orf1ab E1048Q
3474 orf1ab del 2
3479 orf1ab A1072S
4676 orf1ab V1471L
5348 orf1ab silent
5838 orf1ab E1858A
5880 orf1ab S1872I
5907 orf1ab T1881I
6804 orf1ab C2180F
6931 orf1ab del 2
9118 orf1ab silent
9141 orf1ab G2959V
9867 orf1ab L3201P
11288 orf1ab del 9
11601 orf1ab I3779T
12499 orf1ab silent
13242 orf1ab C4326F
13271 orf1ab H4336Y
13337 orf1ab del 1
13339 orf1ab N4358Y
13746 orf1ab silent
14408 orf1ab P314L
14973 orf1ab K502N
15138 orf1ab M557I
15586 orf1ab A707T
15769 orf1ab A768S
16852 orf1ab G1129C
16945 orf1ab A1160S
17565 orf1ab del 1
17721 orf1ab silent
17849 orf1ab G1461V
18344 orf1ab G1626V
18520 orf1ab del 1
18526 orf1ab R1687C
18697 orf1ab del 1
19017 orf1ab silent
19478 orf1ab G2004V
19704 orf1ab silent
19883 orf1ab Y2139S
20262 orf1ab silent
20703 orf1ab del 1
21109 orf1ab T2548A
21451 orf1ab G2662C
21575 S L5F
21846 S T95I
22320 S D253G
22452 S del 1
22508 S del 2
23162 S V534F
23202 S T547R
23294 S D578Y
23302 S Q580H
23387 S A609S
23403 S D614G
23664 S A701V
24504 S del 1
24933 S G1124V
25456 orf3a D22Y
25517 orf3a P42L
25563 orf3a Q57H
26028 orf3a silent
26206 orf3a S272G
26409 E del 1
27115 M R198L
27925 orf8 T11I
28144 orf8 S84L
28271 intergenic del 1
28328 N G19stop
VSP1089−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

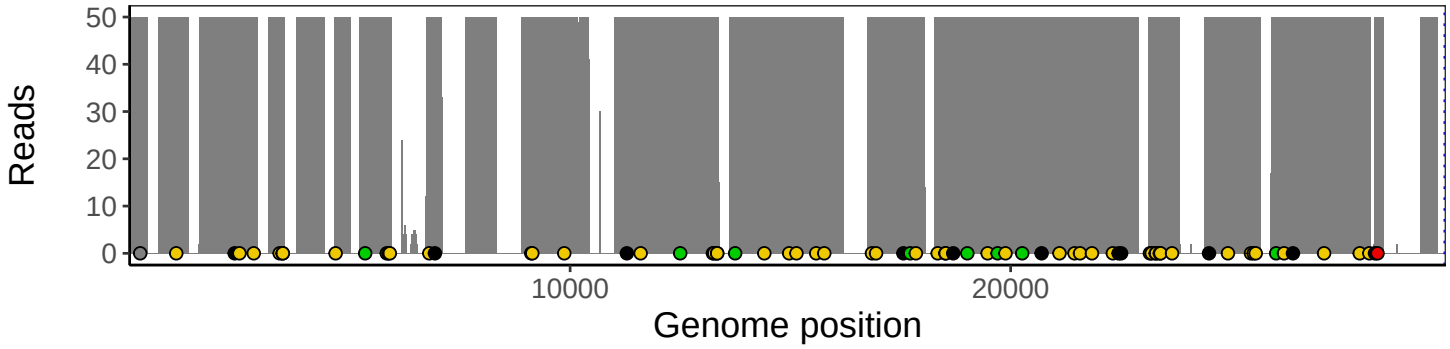
Analyses of individual experiments and composite results

VSP1089-1 | 2021-03-15 | Saline | UPHS-0104 | genomes | single experiment

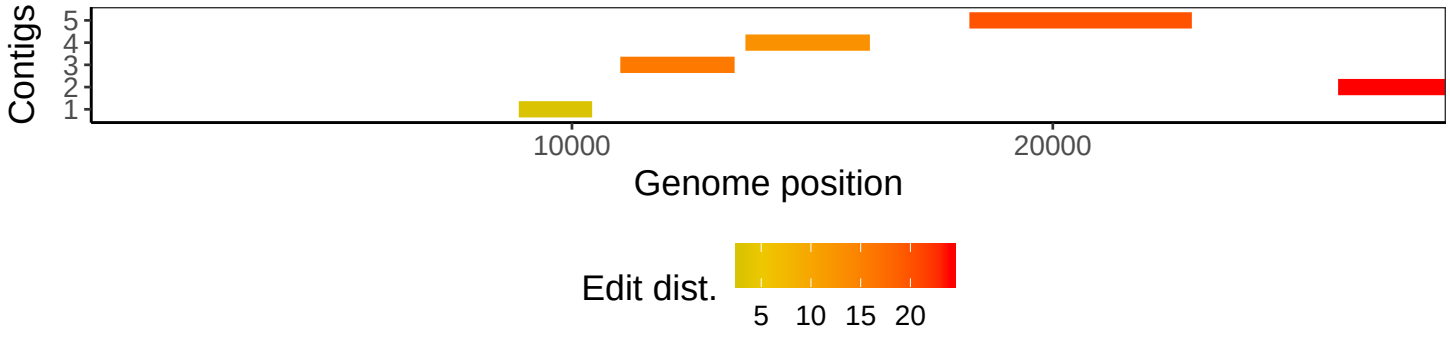
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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