COVID-19 subject UPHS-0815

2021-05-21

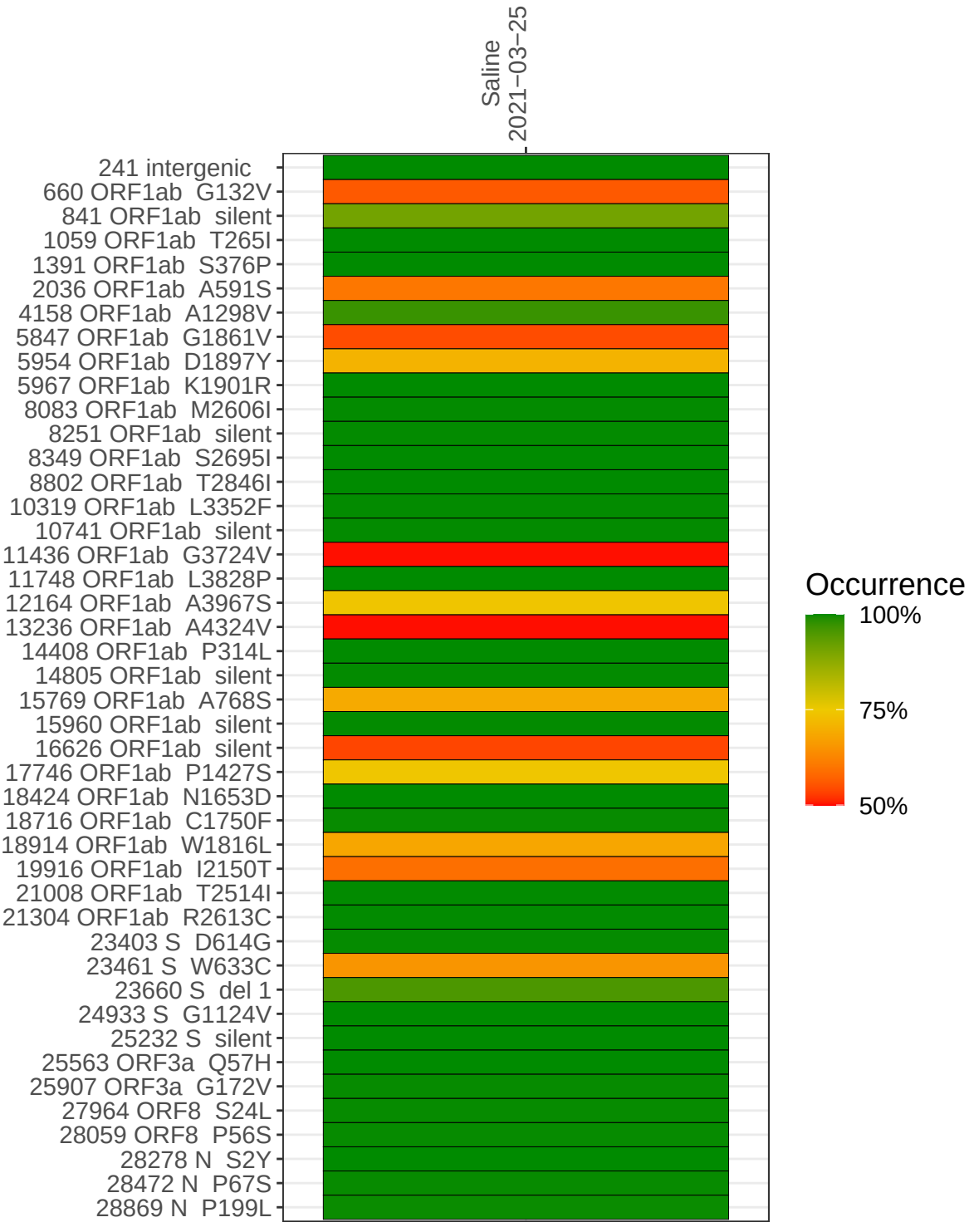
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

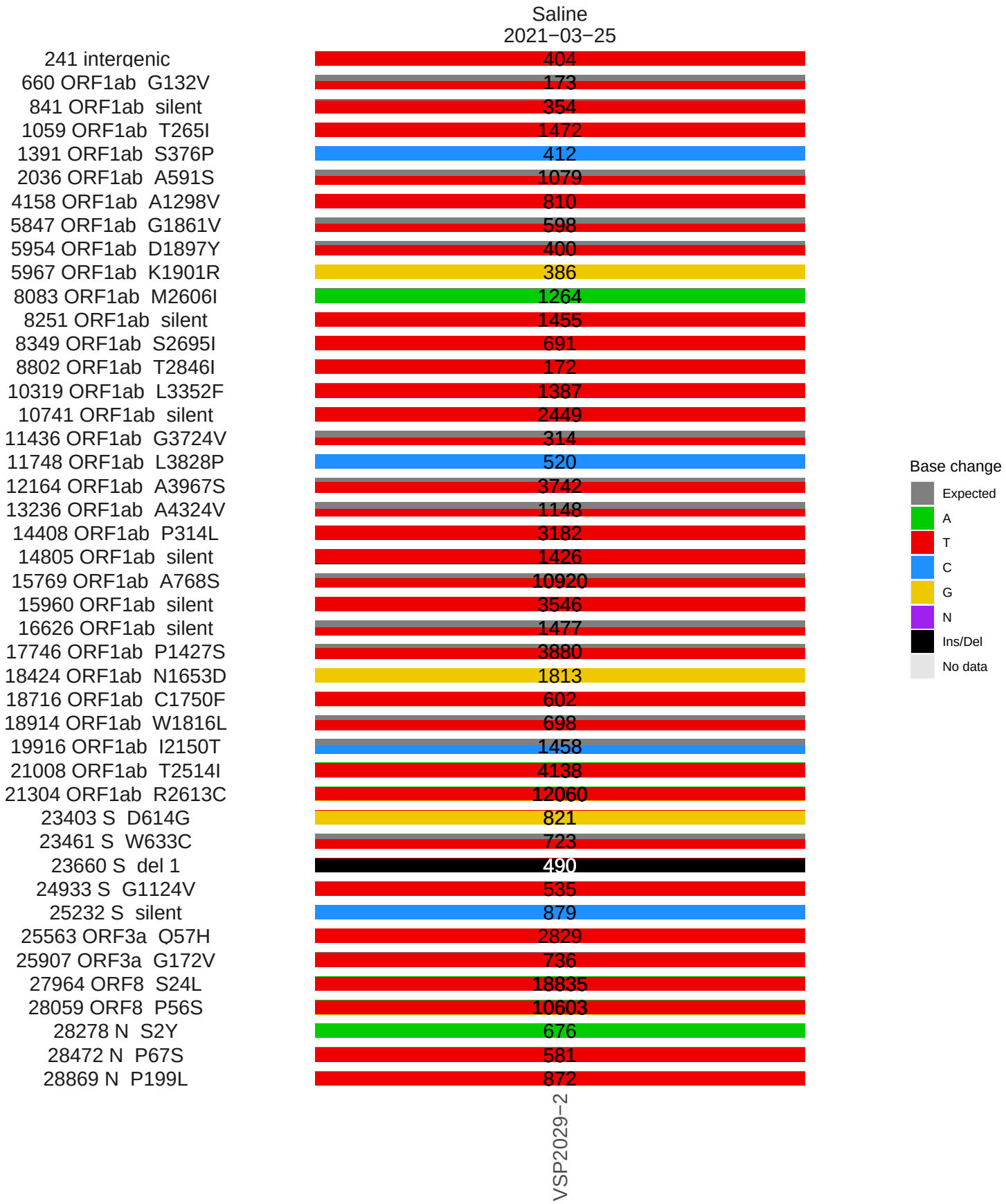
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP2029-2	single experiment	NA	Saline	2021-03-25	9.05	NA	91.3%	91.0%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/Wuhan-Hu-1) 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.

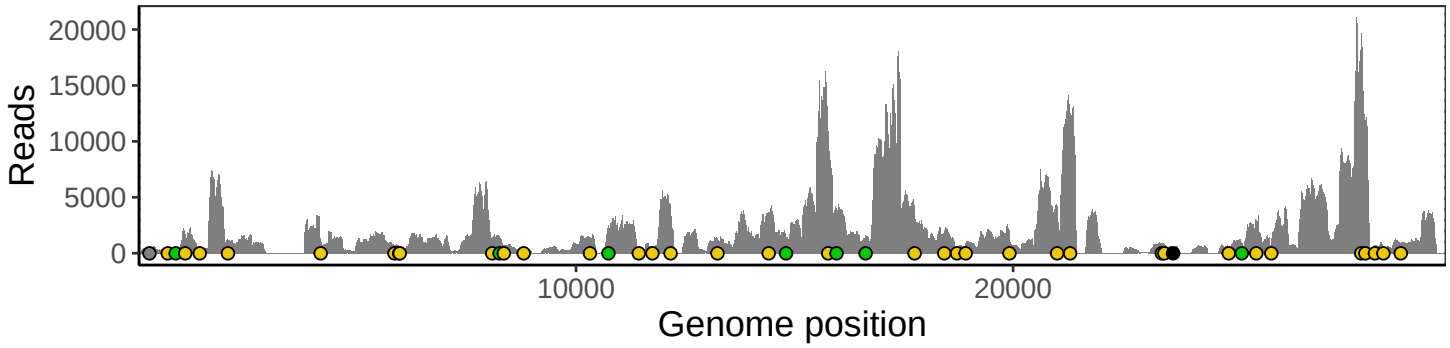



Saline
2021−03−25
241 intergenic 404
660 ORF1ab G132V 173
841 ORF1ab silent 354
1059 ORF1ab T265I 1472
1391 ORF1ab S376P 412
2036 ORF1ab A591S 1079
4158 ORF1ab A1298V 810
5847 ORF1ab G1861V 598
5954 ORF1ab D1897Y 400
5967 ORF1ab K1901R 386
8083 ORF1ab M2606I 1264
8251 ORF1ab silent 1455
8349 ORF1ab S2695I 691
8802 ORF1ab T2846I 172
10319 ORF1ab L3352F 1387
10741 ORF1ab silent 2449
11436 ORF1ab G3724V 314
11748 ORF1ab L3828P 520
12164 ORF1ab A3967S 3742
13236 ORF1ab A4324V 1148
14408 ORF1ab P314L 3182
14805 ORF1ab silent 1426
15769 ORF1ab A768S 10920
15960 ORF1ab silent 3546
16626 ORF1ab silent 1477
17746 ORF1ab P1427S 3880
18424 ORF1ab N1653D 1813
18716 ORF1ab C1750F 602
18914 ORF1ab W1816L 698
19916 ORF1ab I2150T 1458
21008 ORF1ab T2514I 4138
21304 ORF1ab R2613C 12060
23403 S D614G 821
23461 S W633C 723
23660 S del 1 490
24933 S G1124V 535
25232 S silent 879
25563 ORF3a Q57H 2829
25907 ORF3a G172V 736
27964 ORF8 S24L 18835
28059 ORF8 P56S 10603
28278 N S2Y 676
28472 N P67S 581
28869 N P199L 872
VSP2029−2
Base change
Expected
A
T
C
G
N
Ins/Del
No data

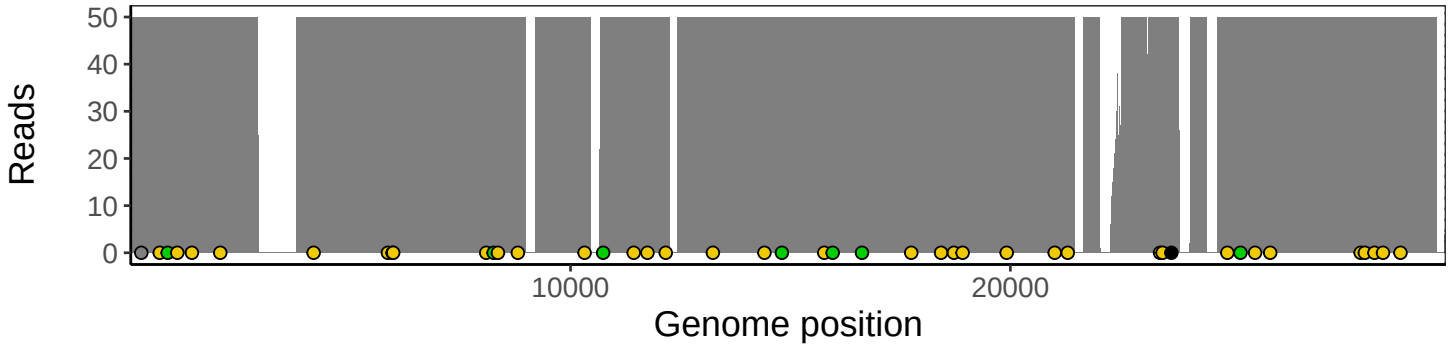
Analyses of individual experiments and composite results

VSP2029-2 | 2021-03-25 | Saline | UPHS-0815 | genomes | single experiment

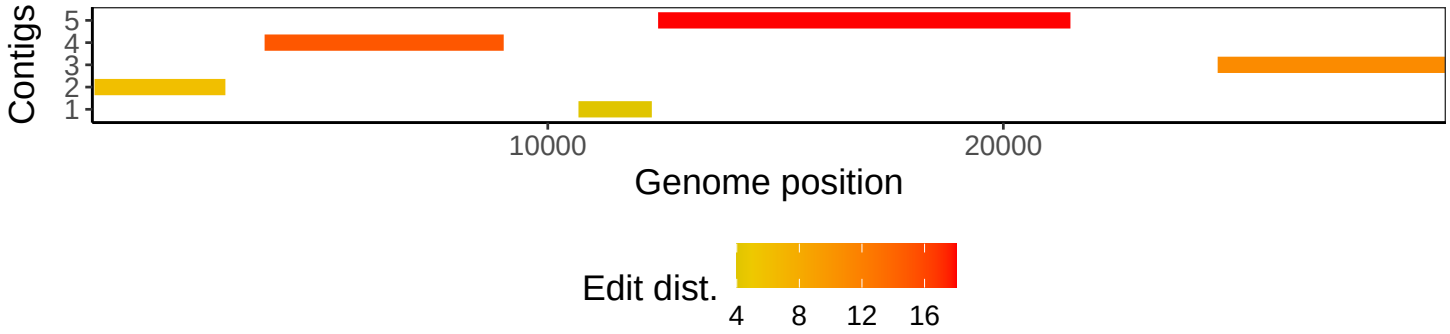
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	3.3.3
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