COVID-19 subject HUP Q-0044

2021-03-29

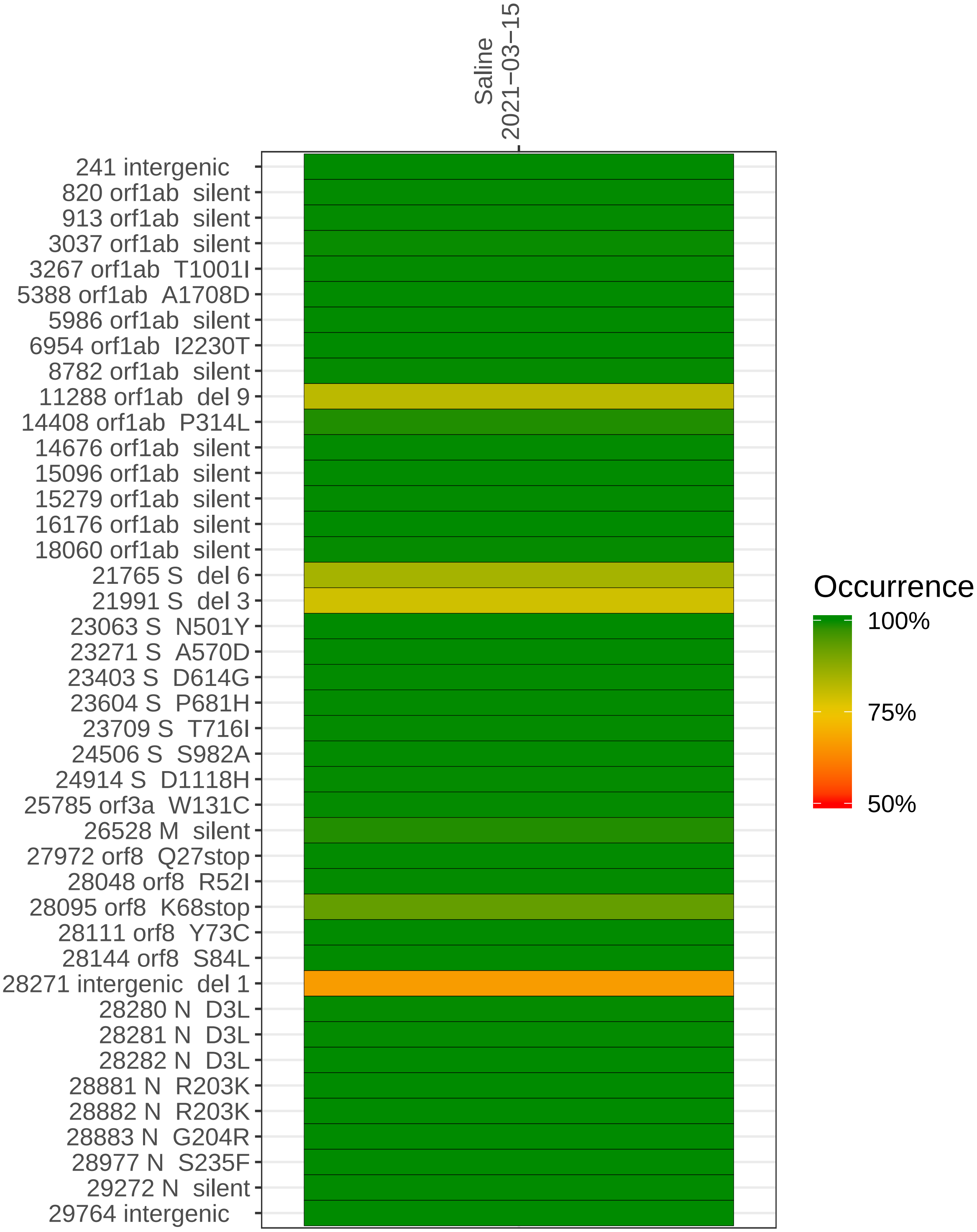
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

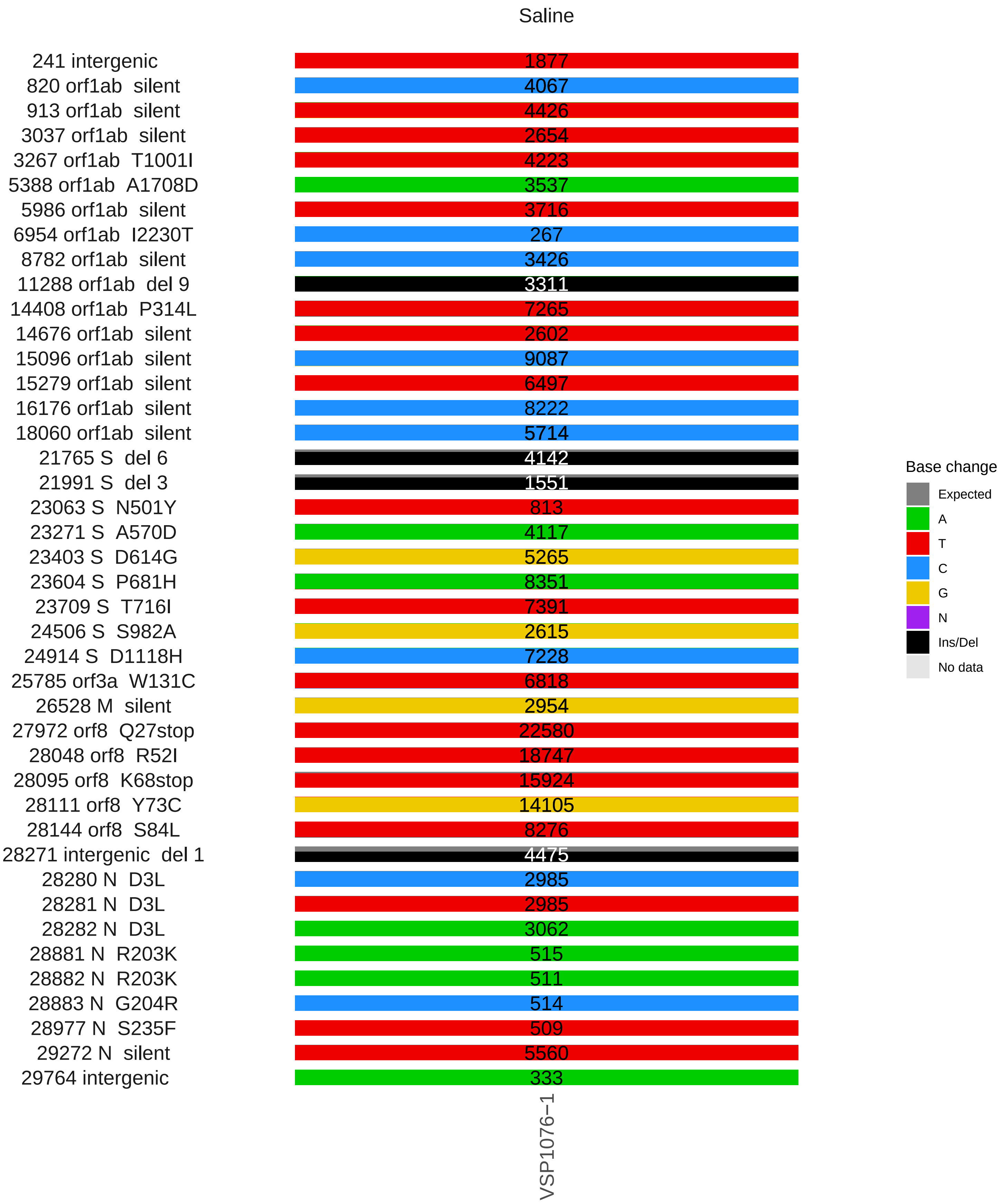
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1076-1	single experiment	NA	Saline	2021-03-15	24.47	B.1.1.7	99.8%	99.8%

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.

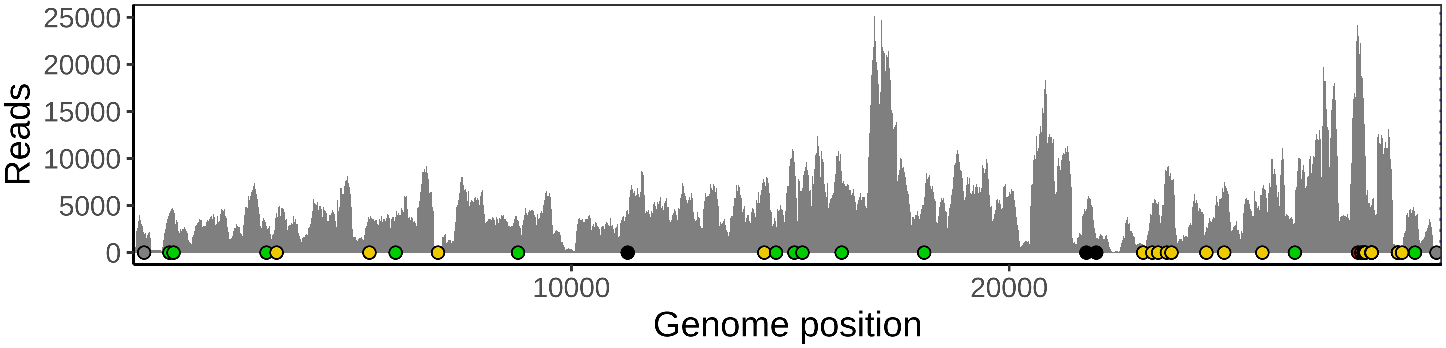



Saline
241 intergenic
1877
820 orf1ab silent
4067
913 orf1ab silent
4426
3037 orf1ab silent
2654
3267 orf1ab T1001I
4223
5388 orf1ab A1708D
3537
5986 orf1ab silent
3716
6954 orf1ab I2230T
267
8782 orf1ab silent
3426
11288 orf1ab del 9
3311
14408 orf1ab P314L
7265
14676 orf1ab silent
2602
15096 orf1ab silent
9087
15279 orf1ab silent
6497
16176 orf1ab silent
8222
18060 orf1ab silent
5714
21765 S del 6
4142
21991 S del 3
1551
23063 S N501Y
813
23271 S A570D
4117
23403 S D614G
5265
23604 S P681H
8351
23709 S T716I
7391
24506 S S982A
2615
24914 S D1118H
7228
25785 orf3a W131C
6818
26528 M silent
2954
27972 orf8 Q27stop
22580
28048 orf8 R52I
18747
28095 orf8 K68stop
15924
28111 orf8 Y73C
14105
28144 orf8 S84L
8276
28271 intergenic del 1
4475
28280 N D3L
2985
28281 N D3L
2985
28282 N D3L
3062
28881 N R203K
515
28882 N R203K
511
28883 N G204R
514
28977 N S235F
509
29272 N silent
5560
29764 intergenic
333
VSP1076–1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

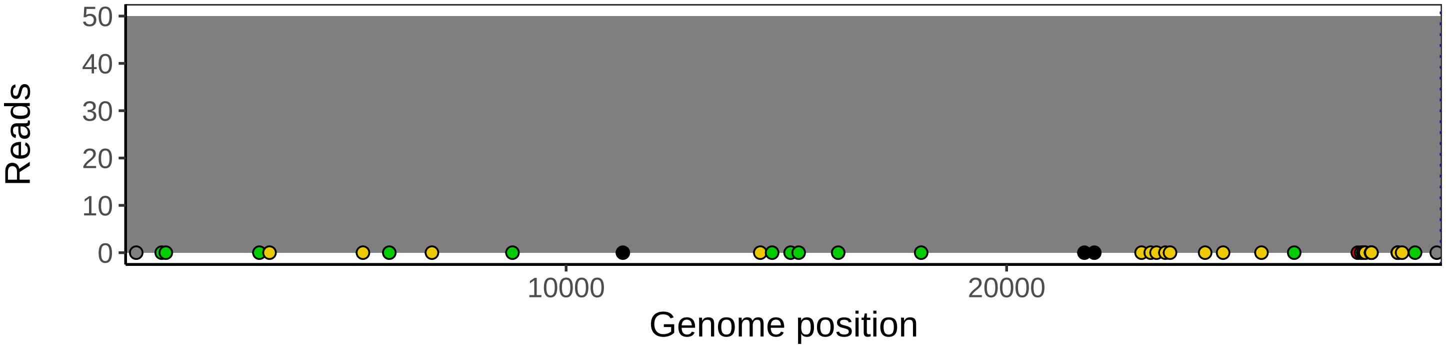
Analyses of individual experiments and composite results

VSP1076-1 | 2021-03-15 | Saline | HUP Q-0044 | genomes | single experiment

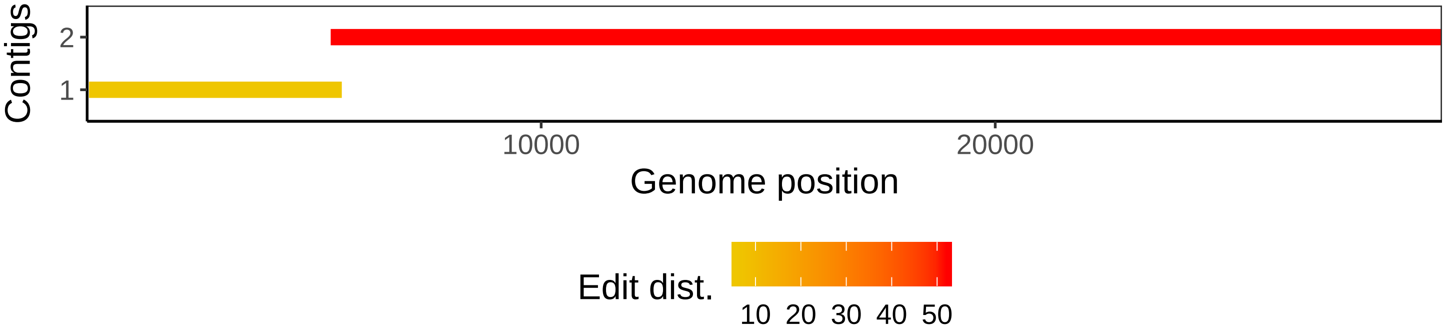
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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