COVID-19 subject MolPath-Seq6

2021-05-05

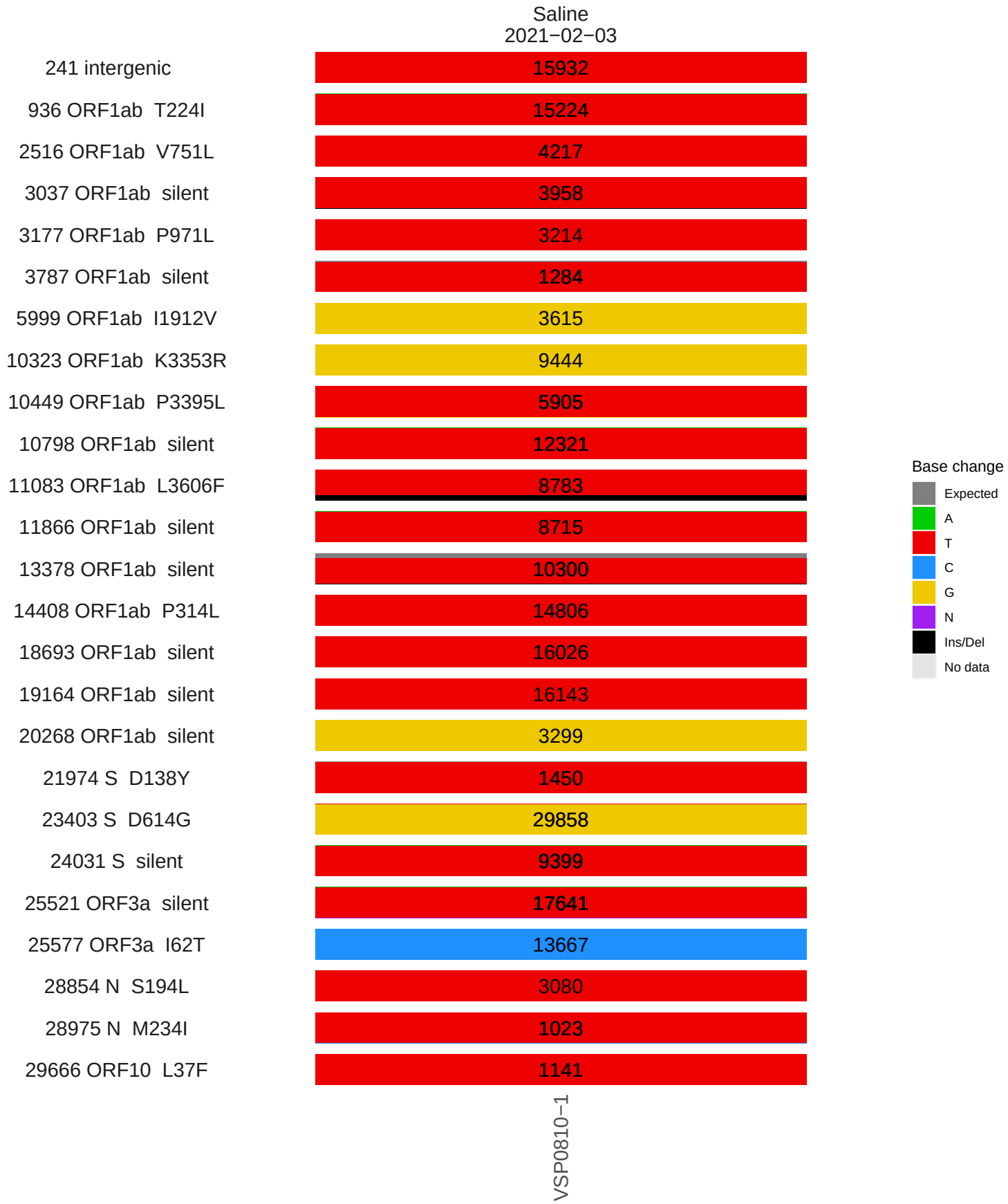
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP0810-1	single experiment	NA	Saline	2021-02-03	29.48	B.1.561	100.0%	99.9%

Variants shared across samples

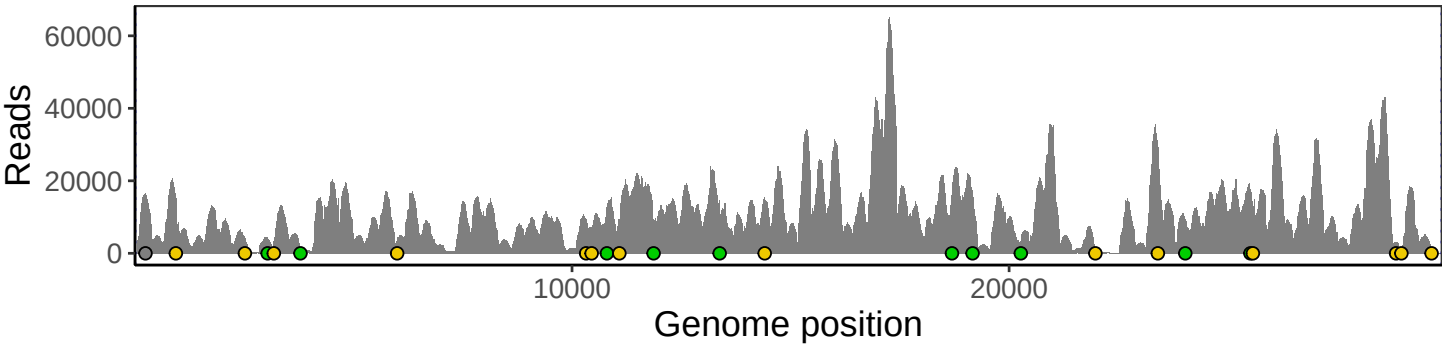
The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/Wuhan-Hu-1) are shared across subject samples where the percent variance is colored. Variants are called if a variant position is covered by 5 or more reads, the alternative base is found in > 50% of 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
2021−02−03
241 intergenic
15932
936 ORF1ab T224I
15224
2516 ORF1ab V751L
4217
3037 ORF1ab silent
3958
3177 ORF1ab P971L
3214
3787 ORF1ab silent
1284
5999 ORF1ab I1912V
3615
10323 ORF1ab K3353R
9444
10449 ORF1ab P3395L
5905
10798 ORF1ab silent
12321
11083 ORF1ab L3606F
8783
11866 ORF1ab silent
8715
13378 ORF1ab silent
10300
14408 ORF1ab P314L
14806
18693 ORF1ab silent
16026
19164 ORF1ab silent
16143
20268 ORF1ab silent
3299
21974 S D138Y
1450
23403 S D614G
29858
24031 S silent
9399
25521 ORF3a silent
17641
25577 ORF3a I62T
13667
28854 N S194L
3080
28975 N M234I
1023
29666 ORF10 L37F
1141
VSP0810−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

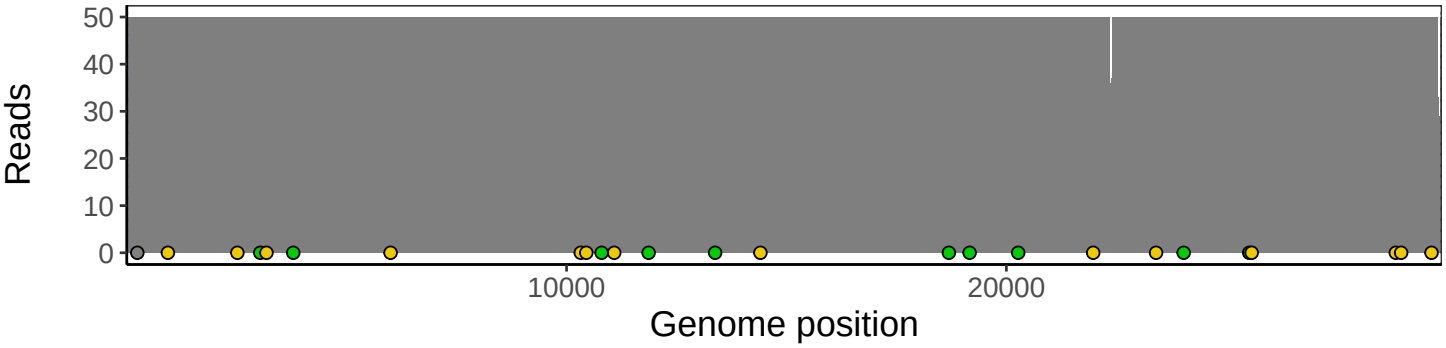
Analyses of individual experiments and composite results

VSP0810-1 | 2021-02-03 | Saline | MolPath-Seq6 | genomes | single experiment

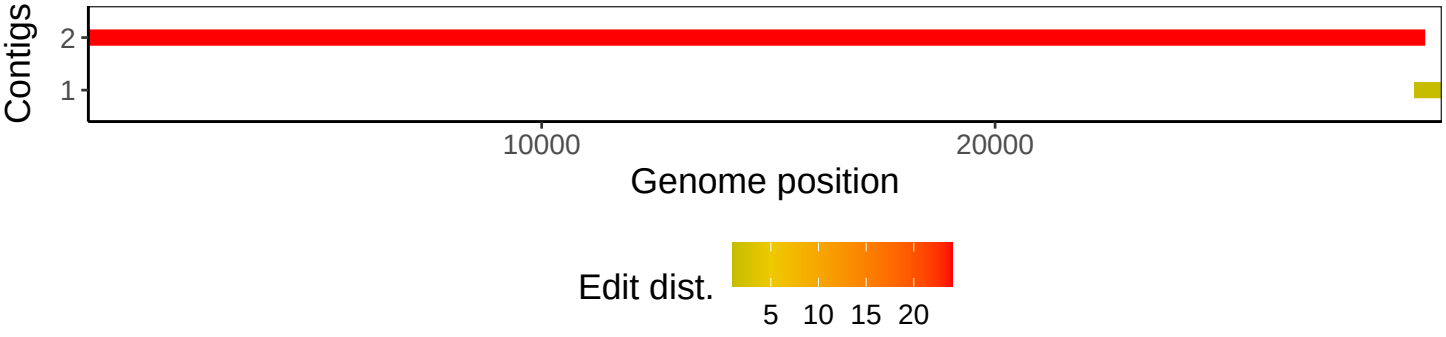
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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