

COVID-19 subject sdrop1__molpath

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.

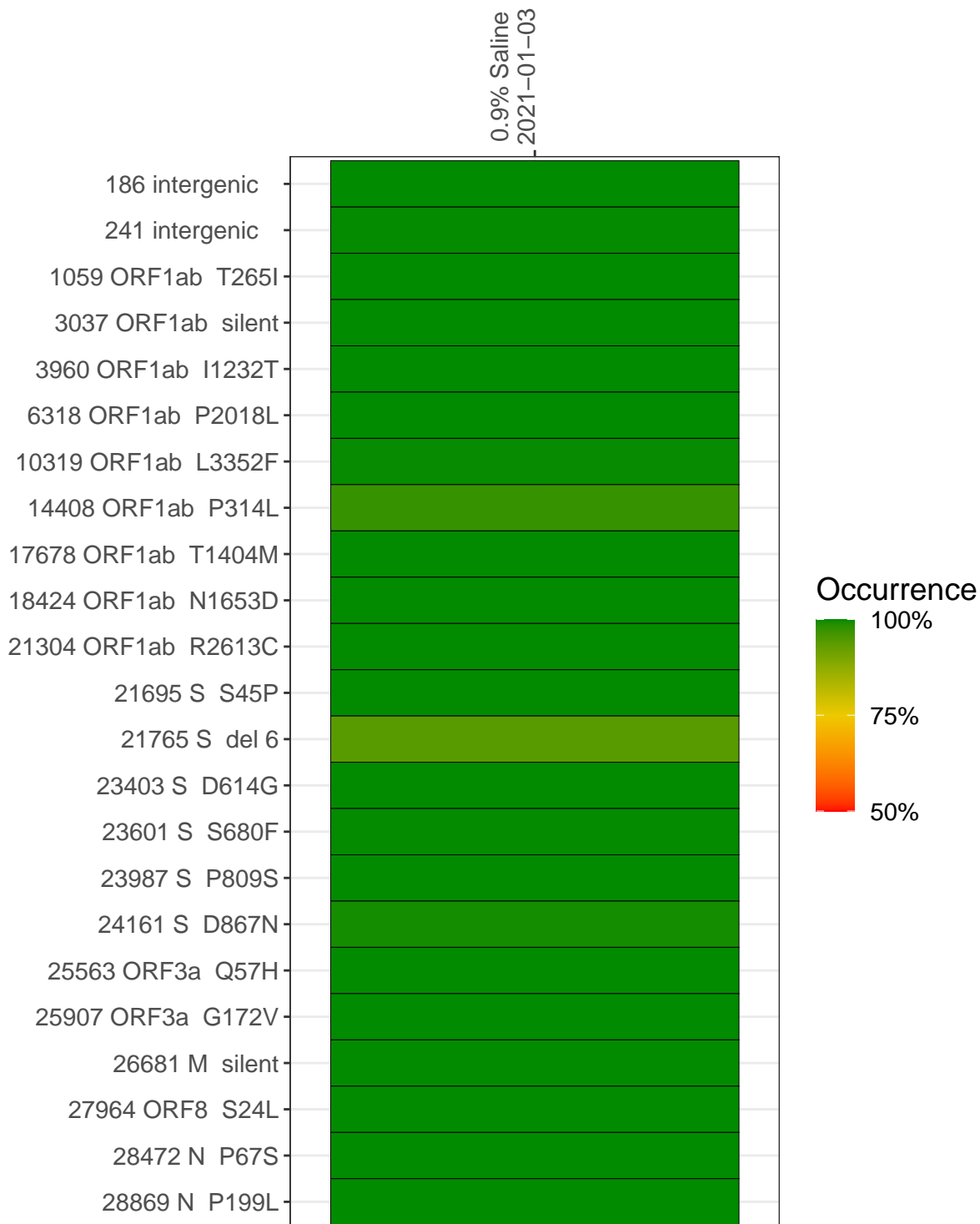
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with $> 90\%$ sequence coverage.

Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (≥ 5 reads)
VSP0571	composite	NA	0.9% Saline	2021-01-03	29.91	B.1.2	99.9%	99.9%
VSP0571-1	single experiment	NA	0.9% Saline	2021-01-03	29.82	B.1.2	99.8%	99.5%
VSP0571-2	single experiment	NA	0.9% Saline	2021-01-03	29.91	B.1.2	99.9%	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.



0.9% Saline
2021-01-03

186 intergenic	1152	22743
241 intergenic	1052	20922
1059 ORF1ab T265I	393	6501
3037 ORF1ab silent	645	9832
3960 ORF1ab I1232T	813	10867
6318 ORF1ab P2018L	1230	19991
10319 ORF1ab L3352F	1073	17584
14408 ORF1ab P314L	1834	25766
17678 ORF1ab T1404M	878	15200
18424 ORF1ab N1653D	1586	27063
21304 ORF1ab R2613C	315	5641
21695 S S45P	242	4341
21765 S del 6	341	6113
23403 S D614G	1473	25711
23601 S S680F	982	16863
23987 S P809S	115	1849
24161 S D867N	378	7115
25563 ORF3a Q57H	632	11420
25907 ORF3a G172V	634	10382
26681 M silent	856	14327
27964 ORF8 S24L	1035	16057
28472 N P67S	1139	20998
28869 N P199L	102	2072

Base change

- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

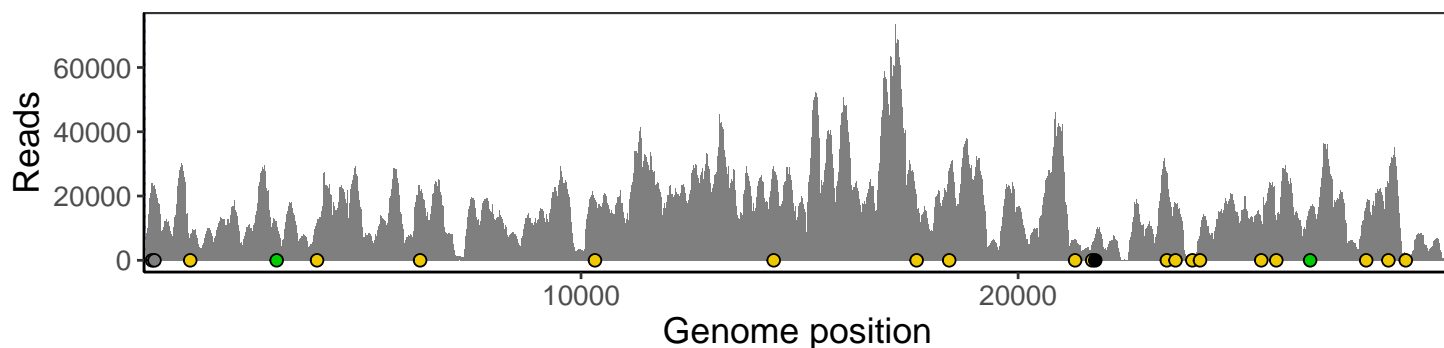
VSP0571-1

VSP0571-2

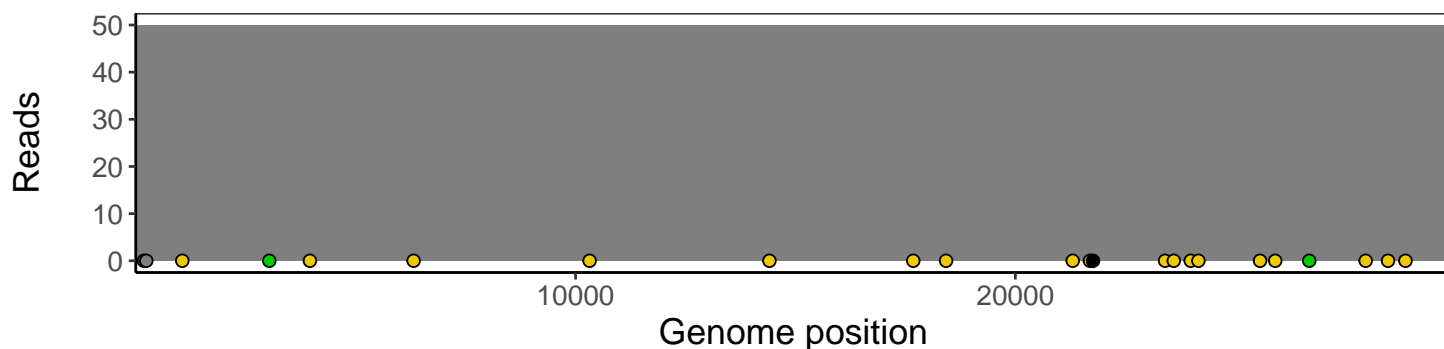
Analyses of individual experiments and composite results

VSP0571 | 2021-01-03 | 0.9% Saline | molpath-sdrop1 | composite result

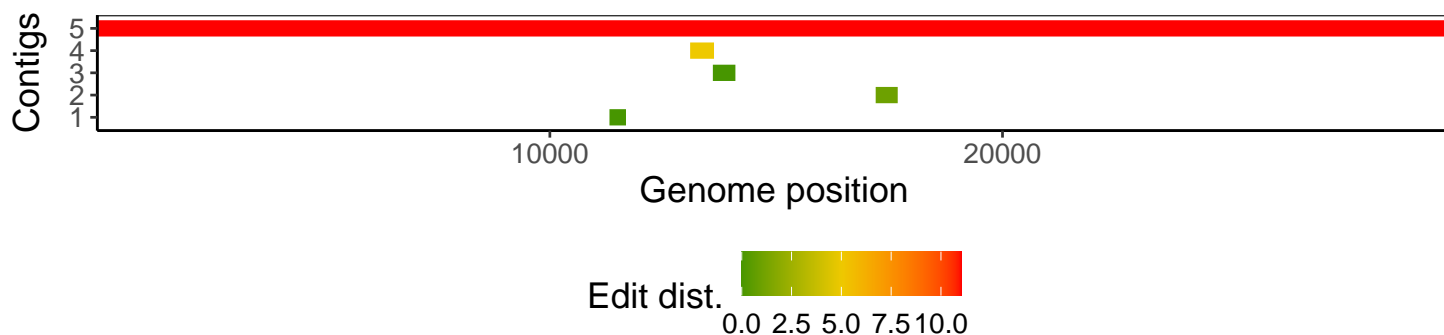
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



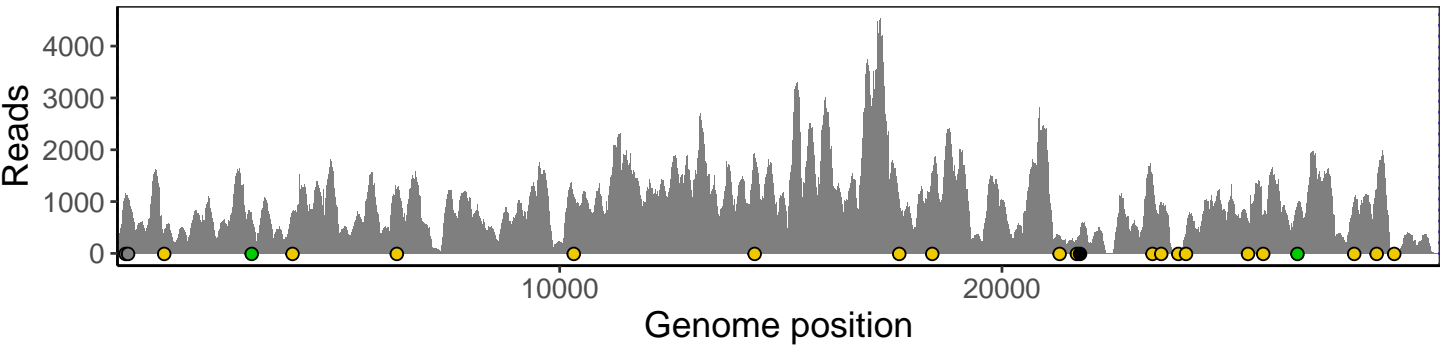
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



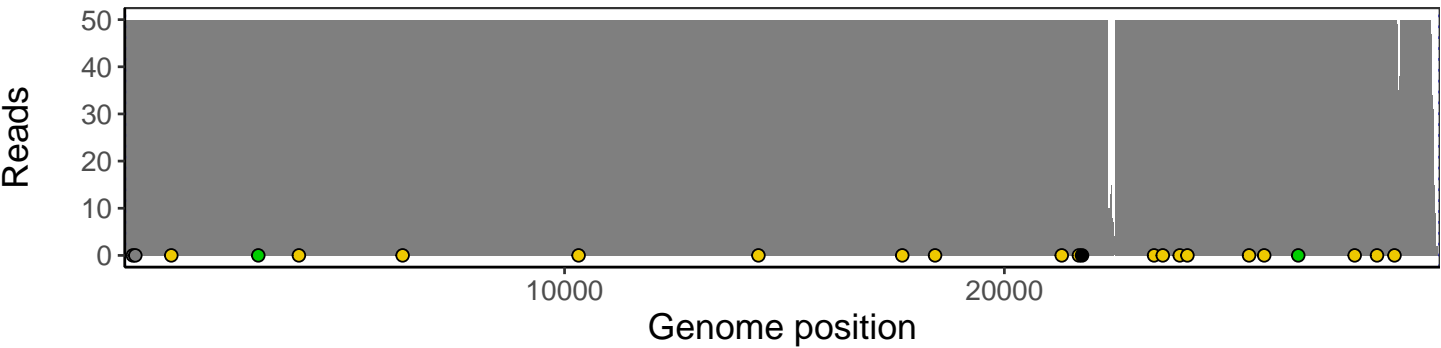
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



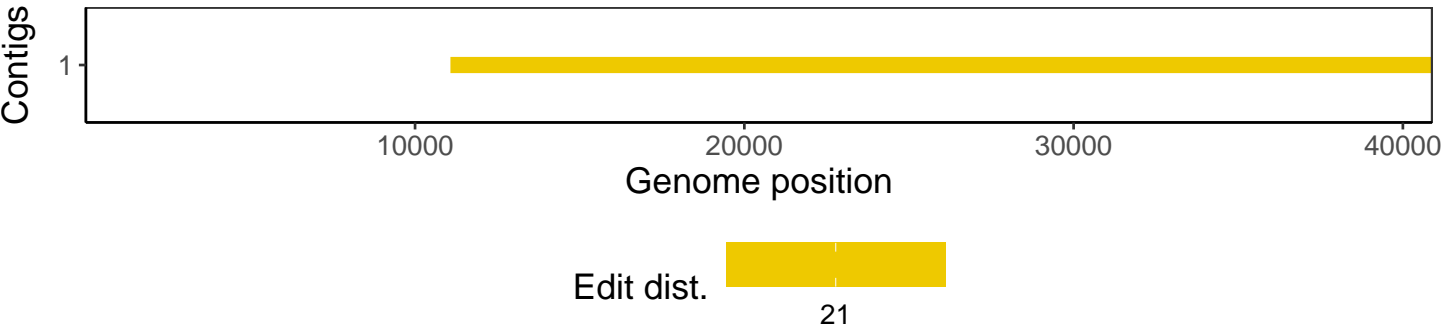
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



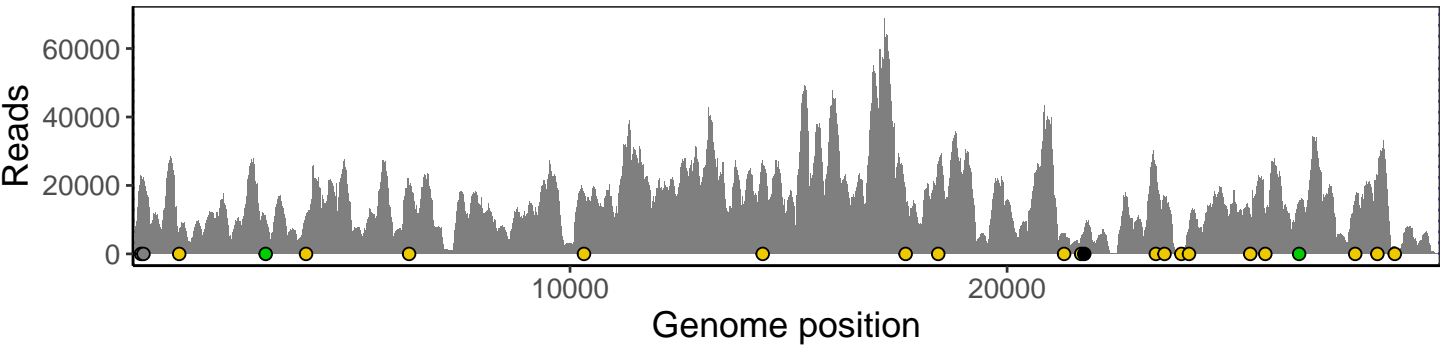
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



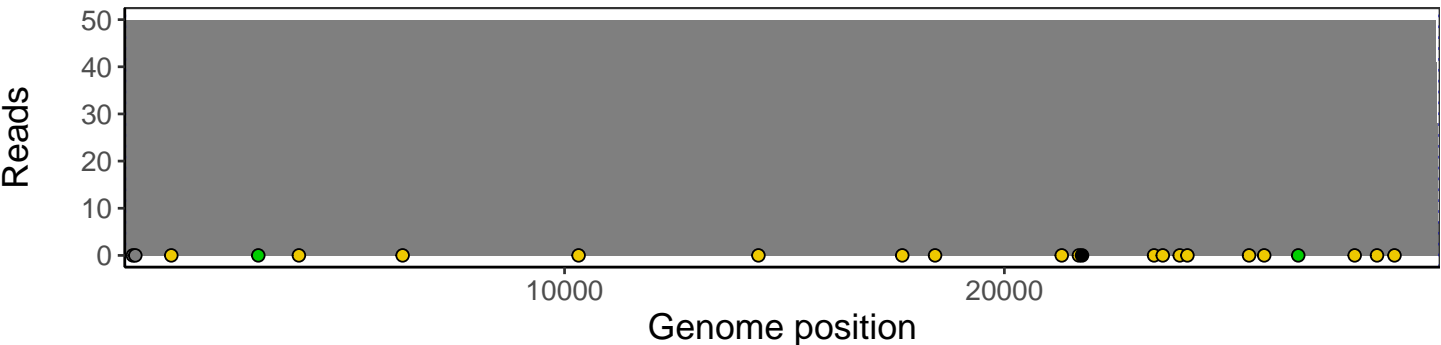
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



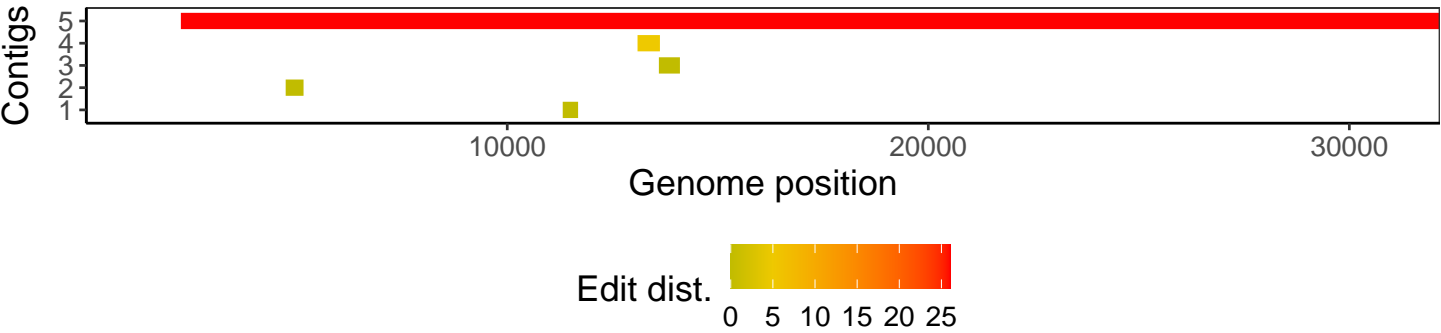
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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 htlib 1.10
bcftools	1.10.2-34-g1a12af0-dirty Using htlib 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