COVID-19 subject sdrop1_molpath

2021-01-31

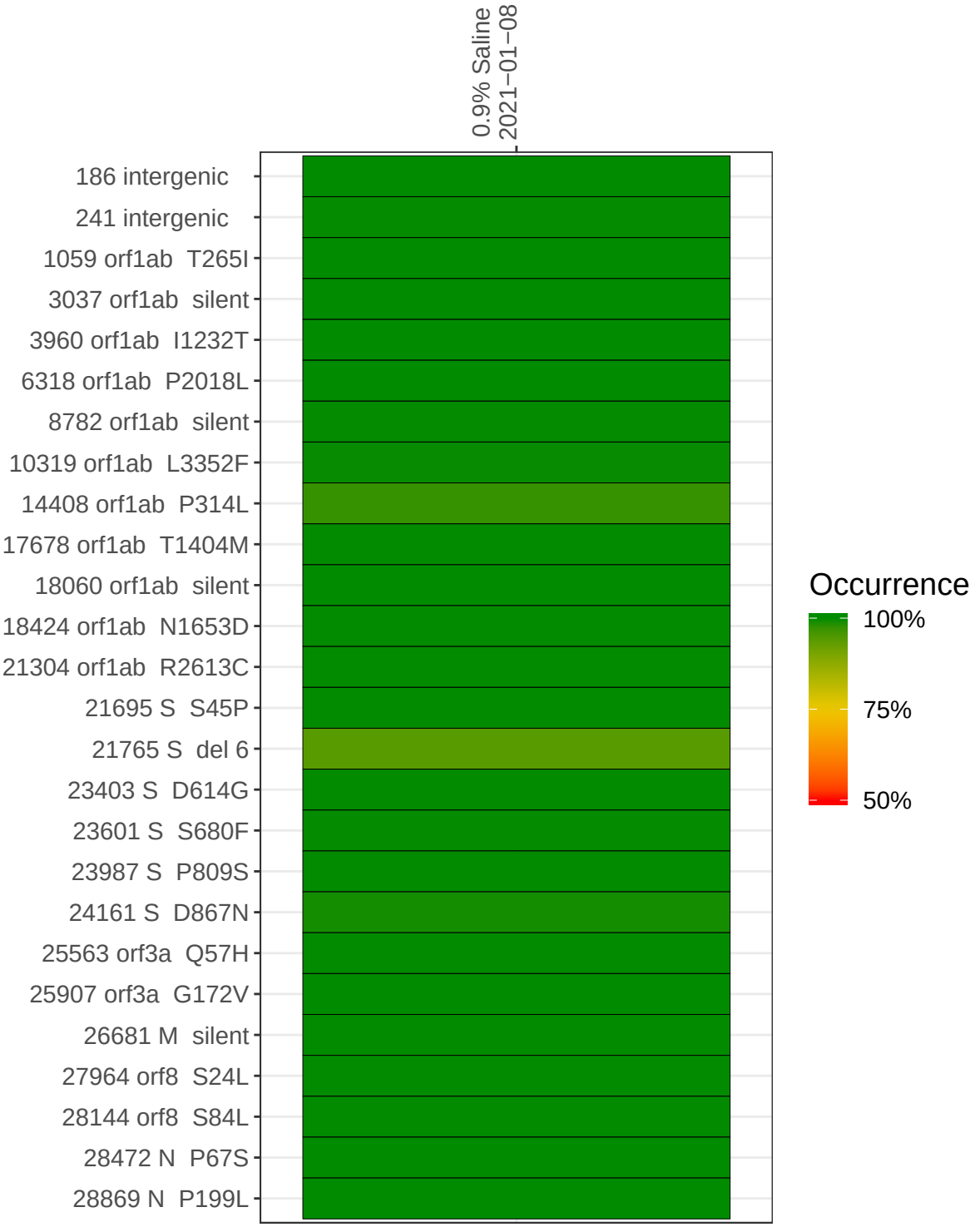
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

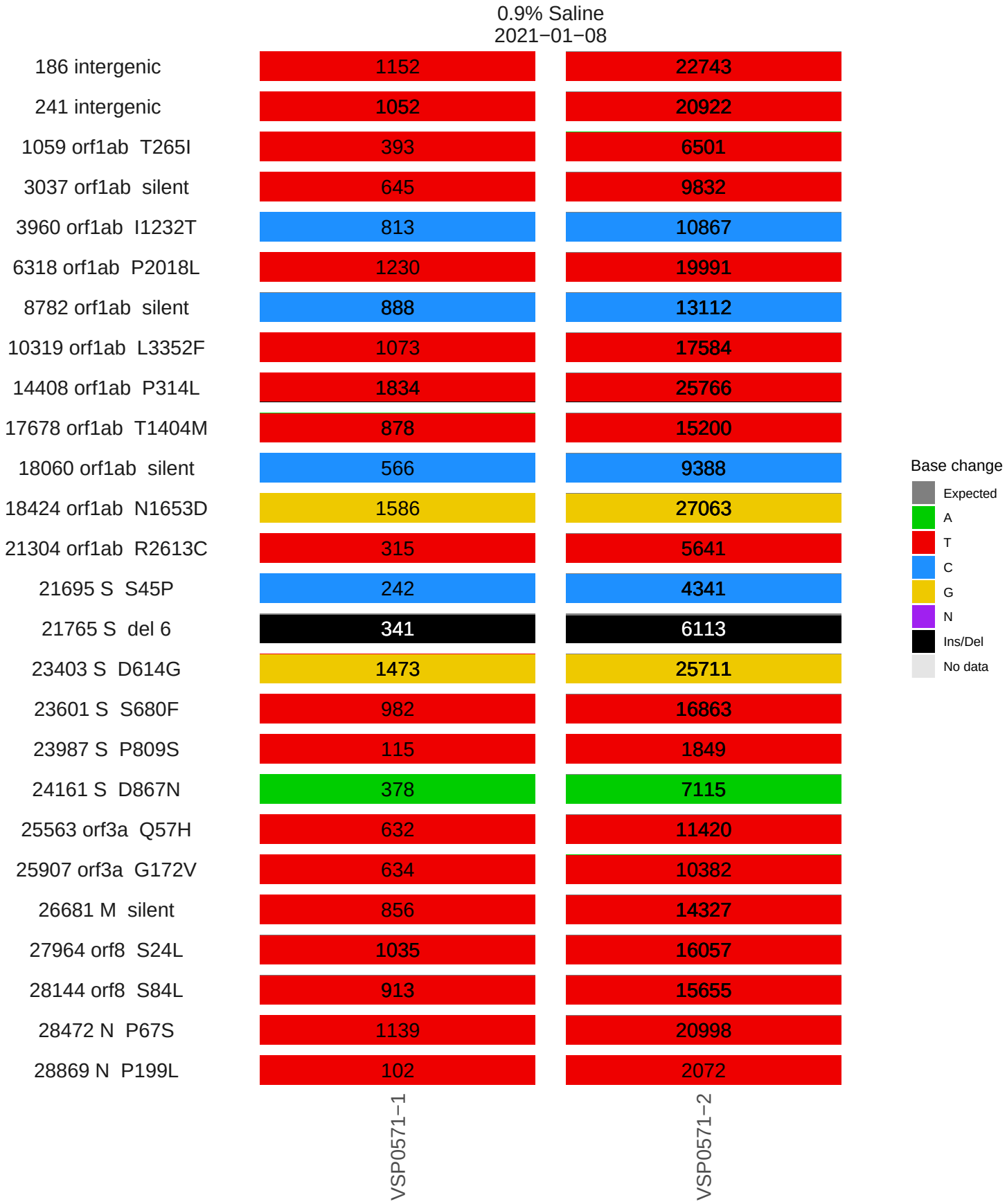
Table 1. Sample summary.

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	2021-01-08	29.91	B.1.2	99.9%	99.9%
VSP0571-1	single experiment	NA	0.9% Saline	2021-01-08	29.82	B.1.2	99.9%	99.6%
VSP0571-2	single experiment	NA	0.9% Saline	2021-01-08	29.91	B.1.2	99.9%	99.9%

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.

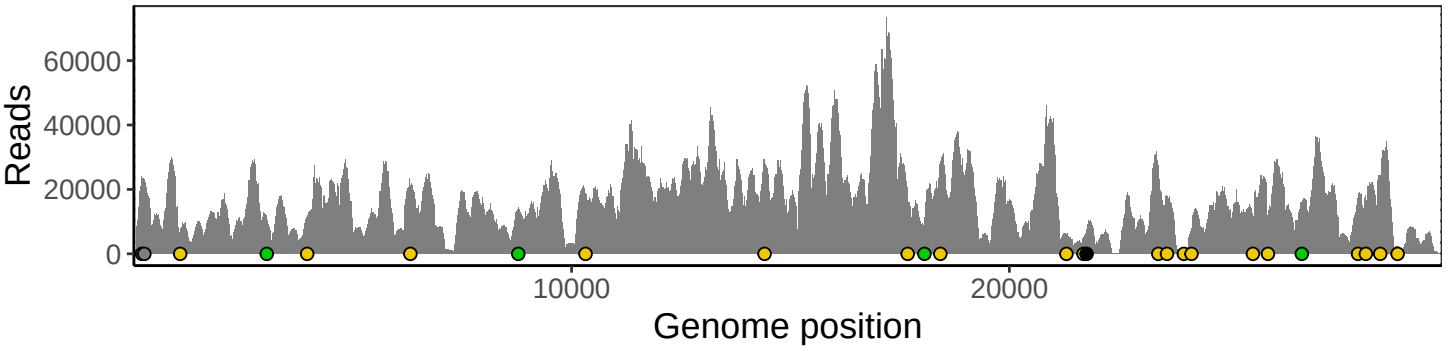



0.9% Saline
2021−01−08
186 intergenic
241 intergenic
1059 orf1ab T265I
3037 orf1ab silent
3960 orf1ab I1232T
6318 orf1ab P2018L
8782 orf1ab silent
10319 orf1ab L3352F
14408 orf1ab P314L
17678 orf1ab T1404M
18060 orf1ab silent
18424 orf1ab N1653D
21304 orf1ab R2613C
21695 S S45P
21765 S del 6
23403 S D614G
23601 S S680F
23987 S P809S
24161 S D867N
25563 orf3a Q57H
25907 orf3a G172V
26681 M silent
27964 orf8 S24L
28144 orf8 S84L
28472 N P67S
28869 N P199L
1152
1052
393
645
813
1230
888
1073
1834
878
566
1586
315
242
341
1473
982
115
378
632
634
856
1035
913
1139
102
22743
20922
6501
9832
10867
19991
13112
17584
25766
15200
9388
27063
5641
4341
6113
25711
16863
1849
7115
11420
10382
14327
16057
15655
20998
2072
VSP0571−1
VSP0571−2
Base change
Expected
A
T
C
G
N
Ins/Del
No data

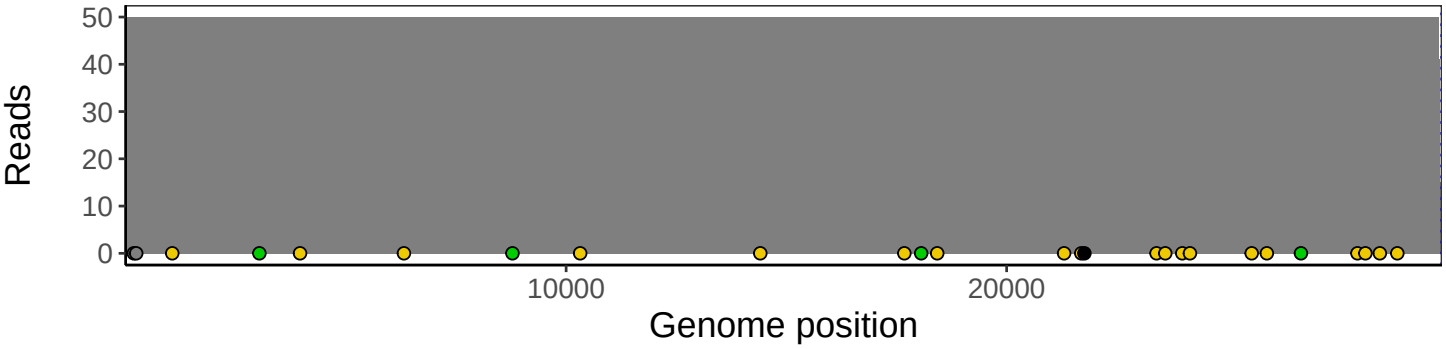
Analyses of individual experiments and composite results

VSP0571 | 2021-01-08 | 0.9% Saline | sdrop1_molpath | composite result

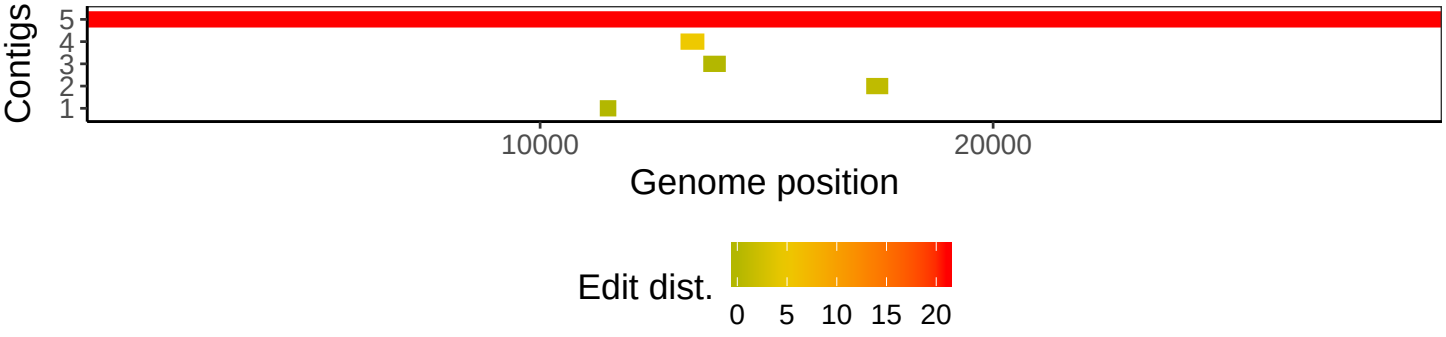
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

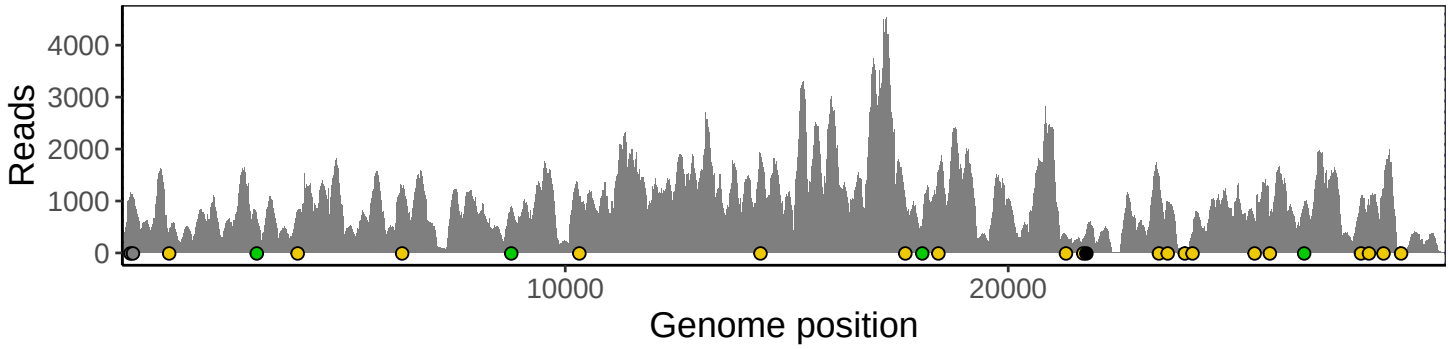


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

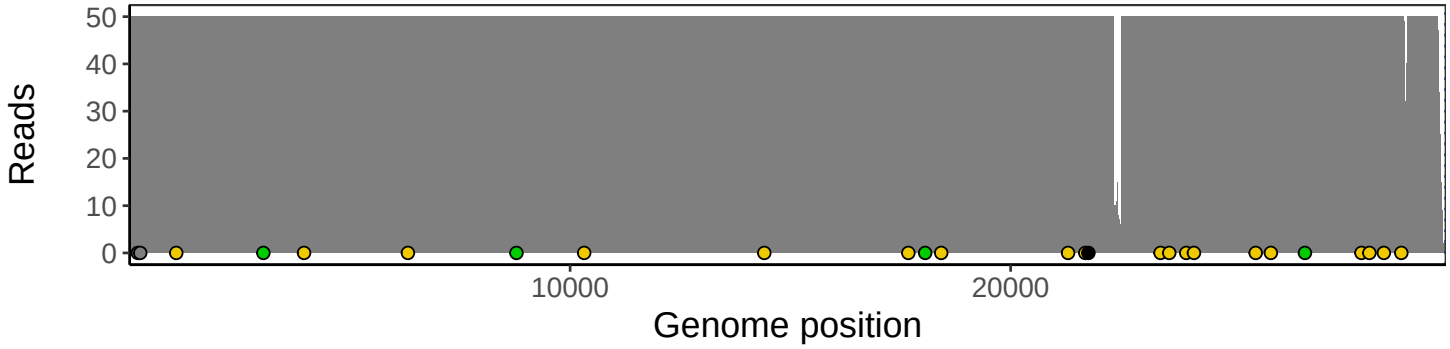


VSP0571-1 | 2021-01-08 | 0.9% Saline | sdrop1_molpath | genomes | single experiment

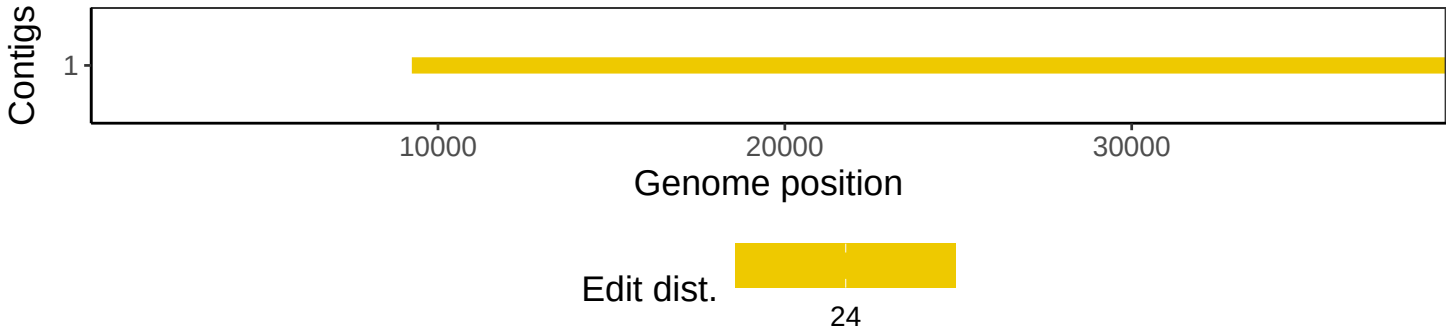
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

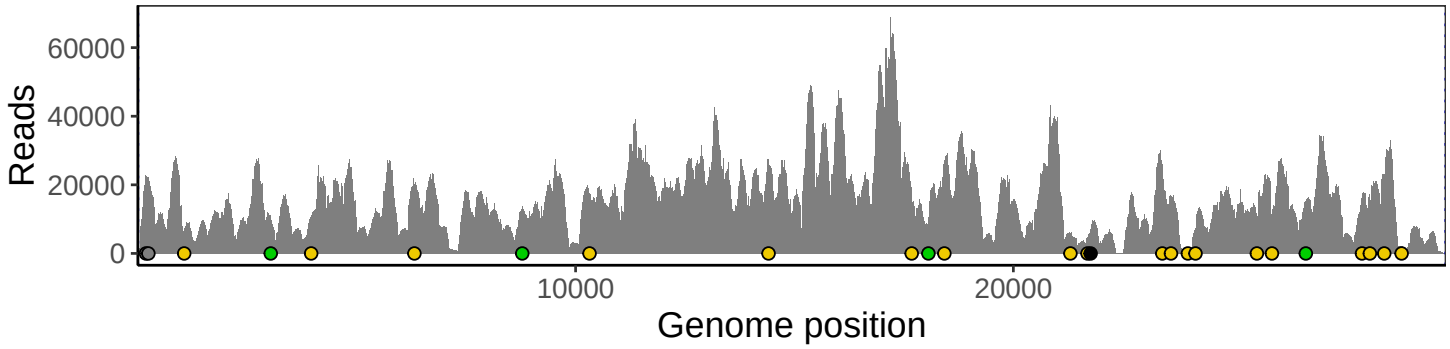


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

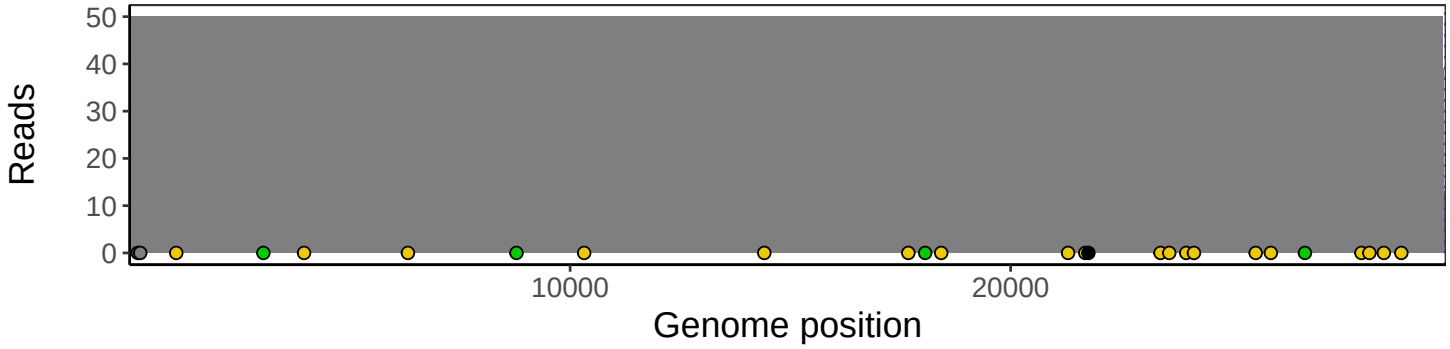


VSP0571-2 | 2021-01-08 | 0.9% Saline | sdrop1_molpath | genomes | single experiment

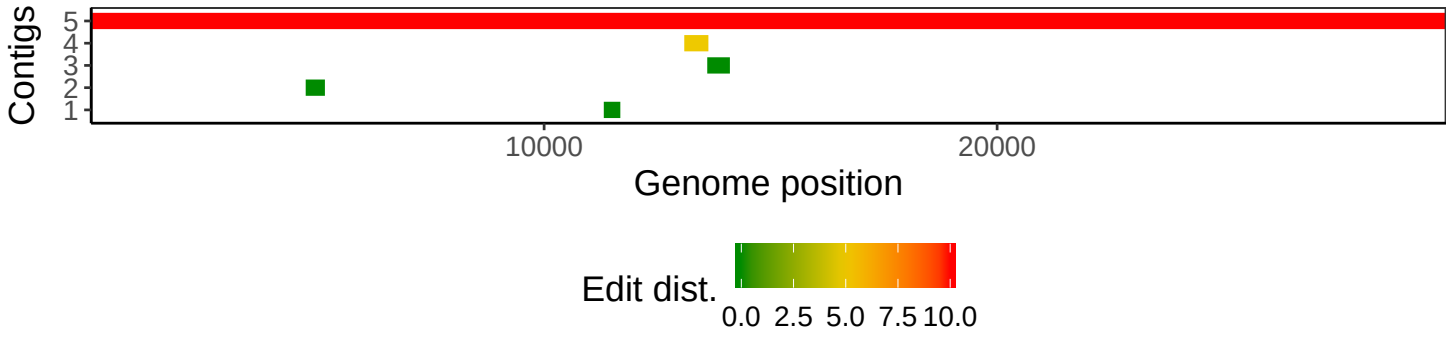
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.1.7
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