COVID-19 subject molpath-seq1

2021-03-01

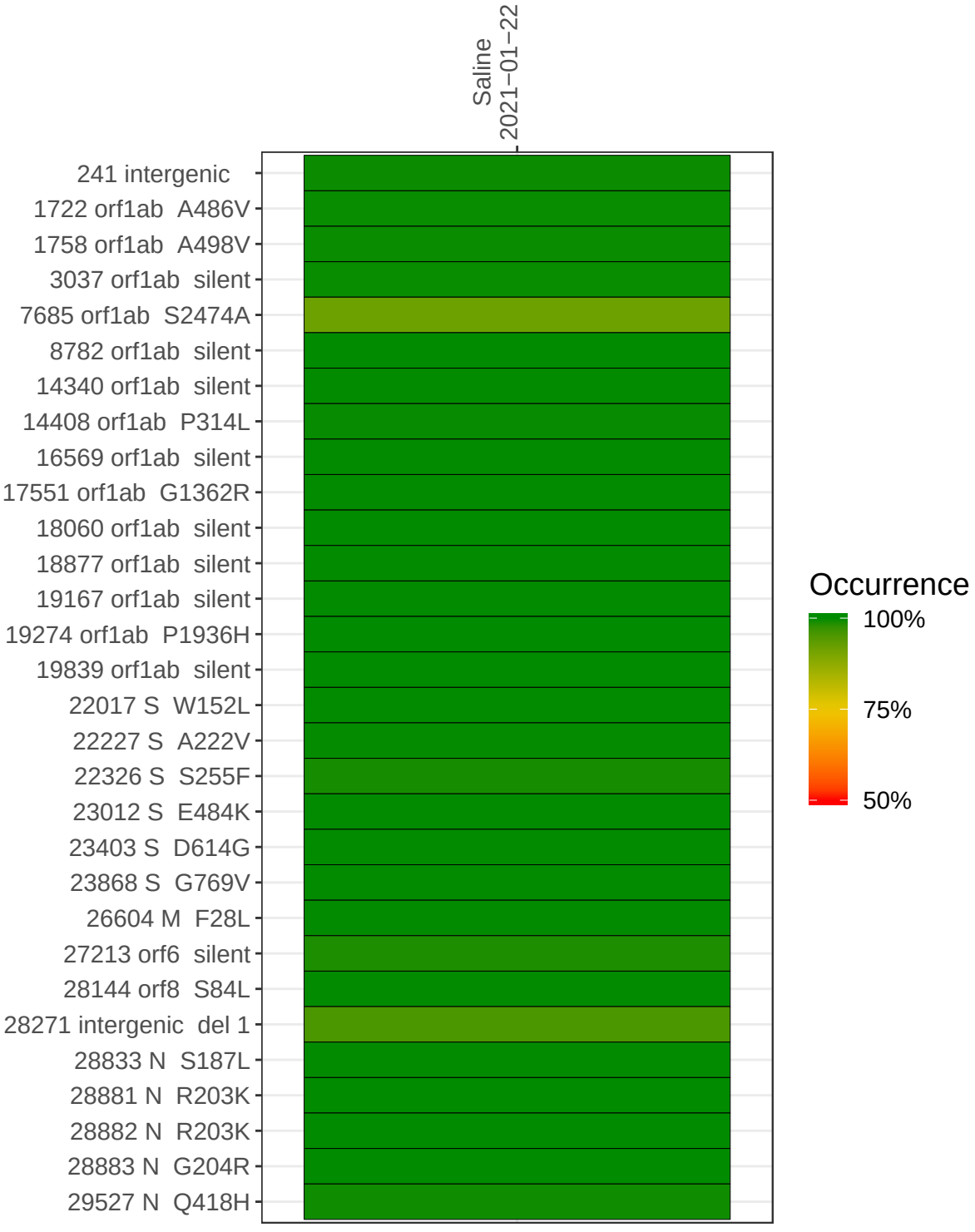
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

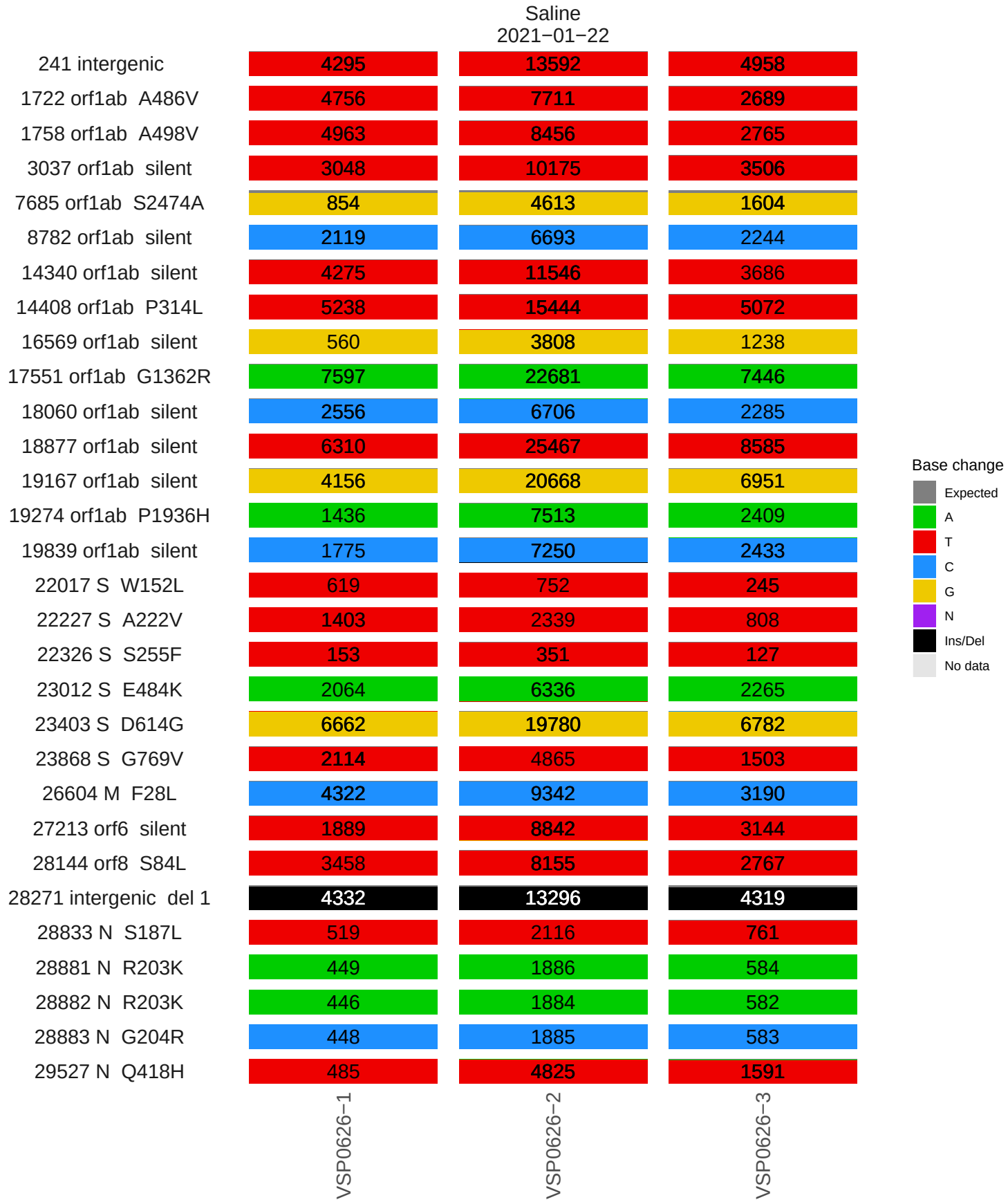
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0626	composite	NA	Saline	2021-01-22	20.39	R.1	99.7%	99.1%
VSP0626-1	single experiment	NA	Saline	2021-01-22	21.48	R.1	99.0%	99.0%
VSP0626-2	single experiment	NA	Saline	2021-01-22	20.65	R.1	99.7%	99.1%
VSP0626-3	single experiment	NA	Saline	2021-01-22	21.80	R.1	99.0%	98.8%

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.

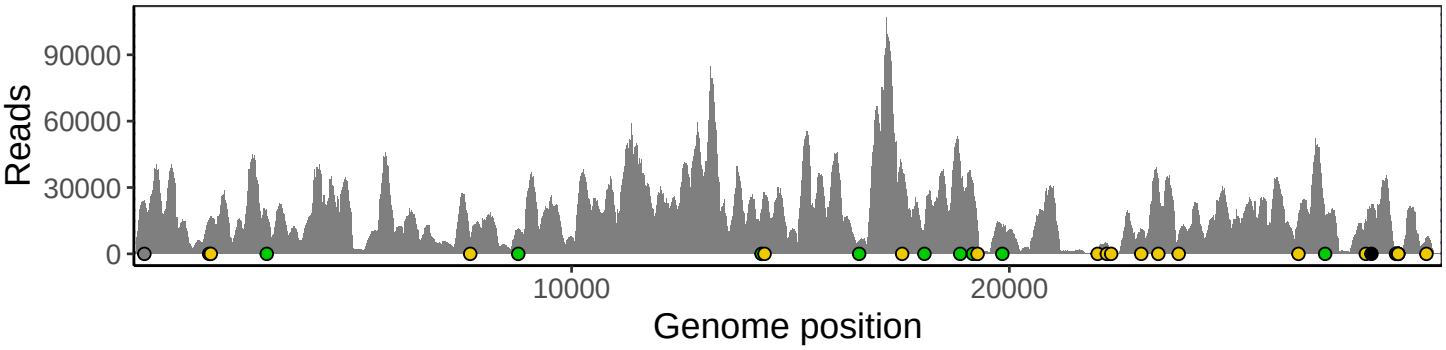



Saline
2021−01−22
241 intergenic
1722 orf1ab A486V
1758 orf1ab A498V
3037 orf1ab silent
7685 orf1ab S2474A
8782 orf1ab silent
14340 orf1ab silent
14408 orf1ab P314L
16569 orf1ab silent
17551 orf1ab G1362R
18060 orf1ab silent
18877 orf1ab silent
19167 orf1ab silent
19274 orf1ab P1936H
19839 orf1ab silent
22017 S W152L
22227 S A222V
22326 S S255F
23012 S E484K
23403 S D614G
23868 S G769V
26604 M F28L
27213 orf6 silent
28144 orf8 S84L
28271 intergenic del 1
28833 N S187L
28881 N R203K
28882 N R203K
28883 N G204R
29527 N Q418H
4295 13592 4958
4756 7711 2689
4963 8456 2765
3048 10175 3506
854 4613 1604
2119 6693 2244
4275 11546 3686
5238 15444 5072
560 3808 1238
7597 22681 7446
2556 6706 2285
6310 25467 8585
4156 20668 6951
1436 7513 2409
1775 7250 2433
619 752 245
1403 2339 808
153 351 127
2064 6336 2265
6662 19780 6782
2114 4865 1503
4322 9342 3190
1889 8842 3144
3458 8155 2767
4332 13296 4319
519 2116 761
449 1886 584
446 1884 582
448 1885 583
485 4825 1591
VSP0626−1
VSP0626−2
VSP0626−3
Base change
Expected
A
T
C
G
N
Ins/Del
No data

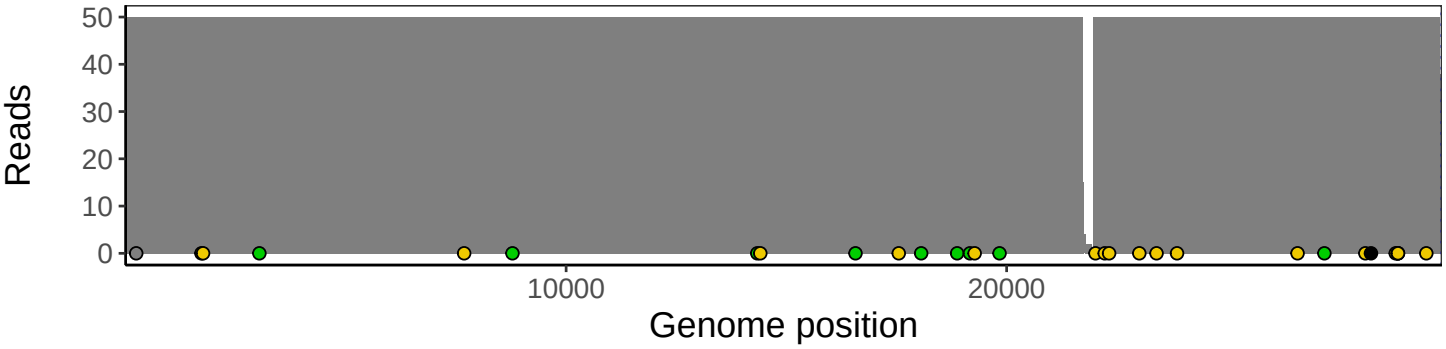
Analyses of individual experiments and composite results

VSP0626 | 2021-01-22 | Saline | molpath-seq1 | composite result

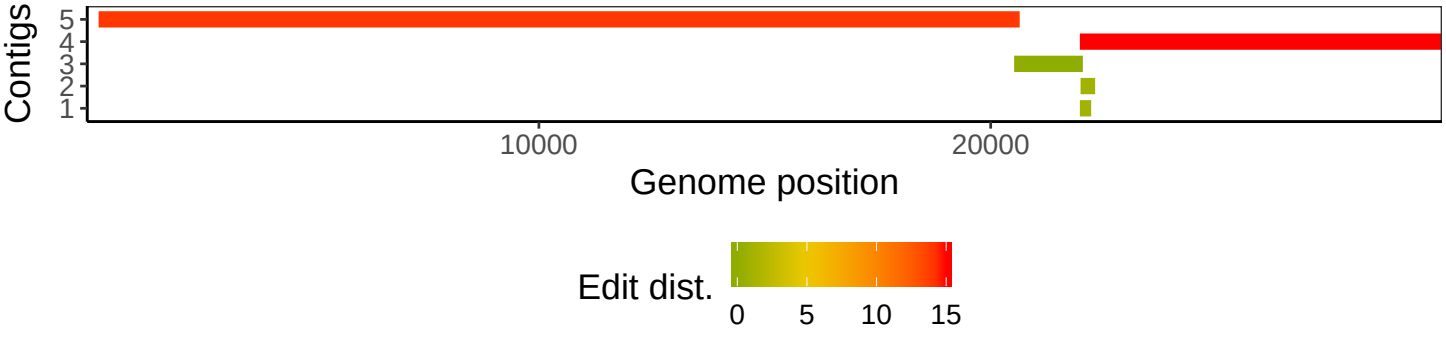
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

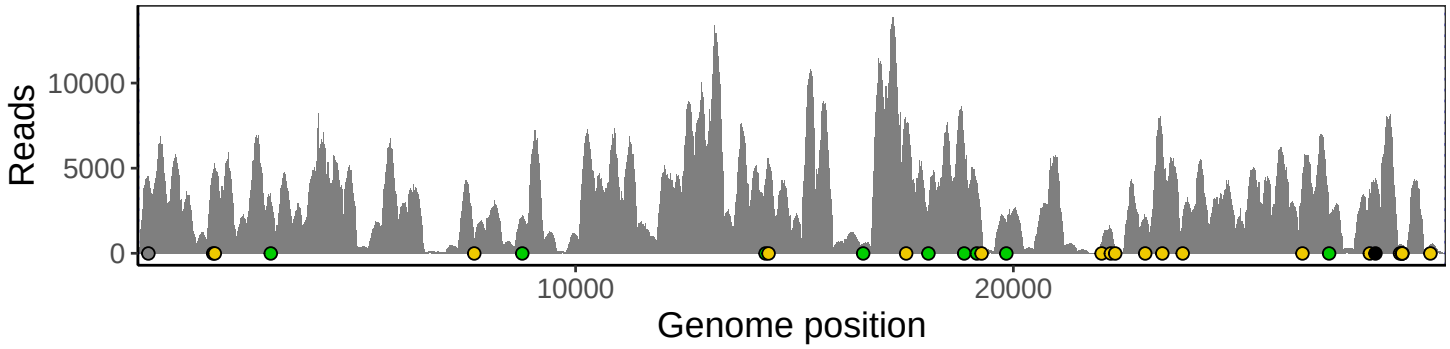


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

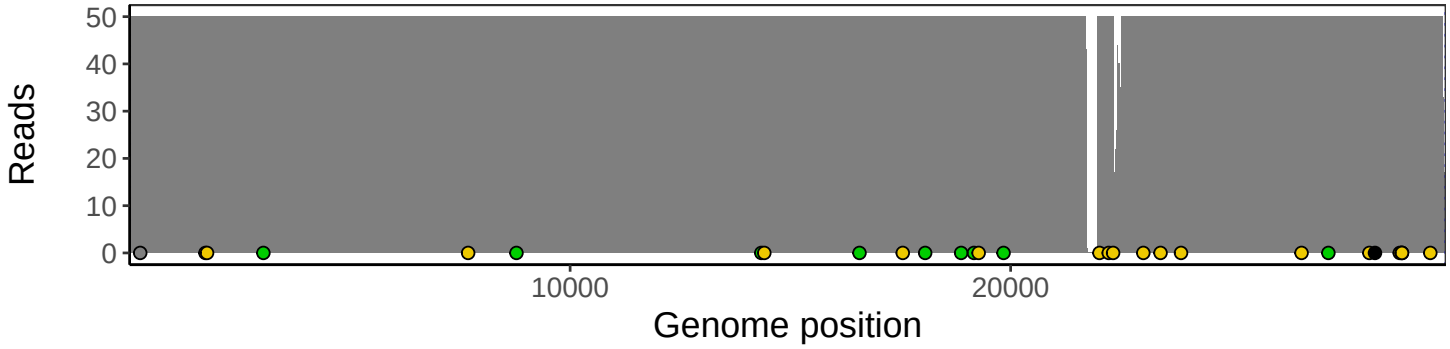


VSP0626-1 | 2021-01-22 | Saline | molpath-seq1 | genomes | single experiment

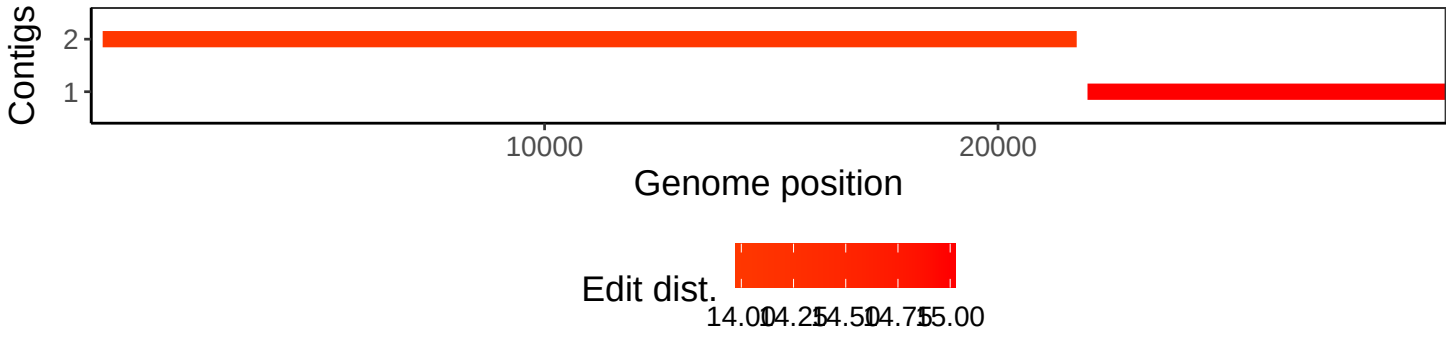
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

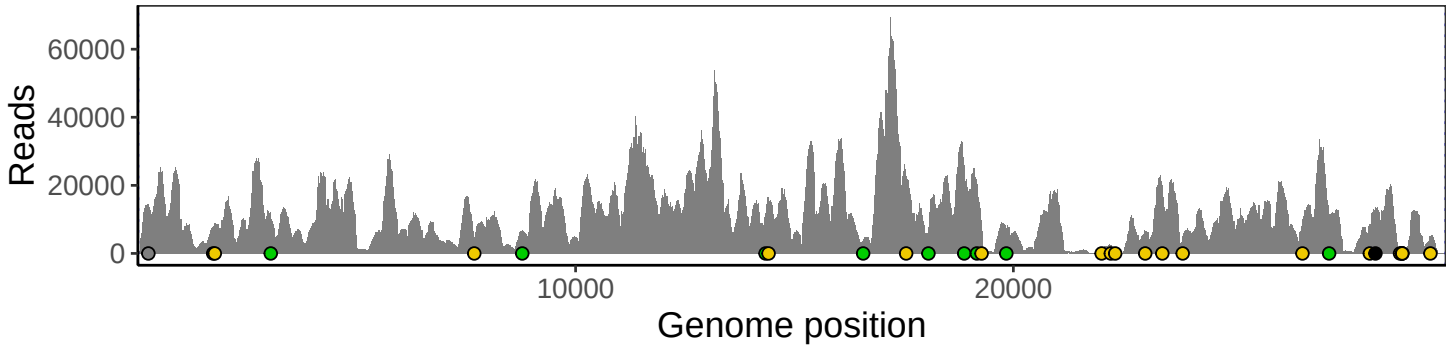


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

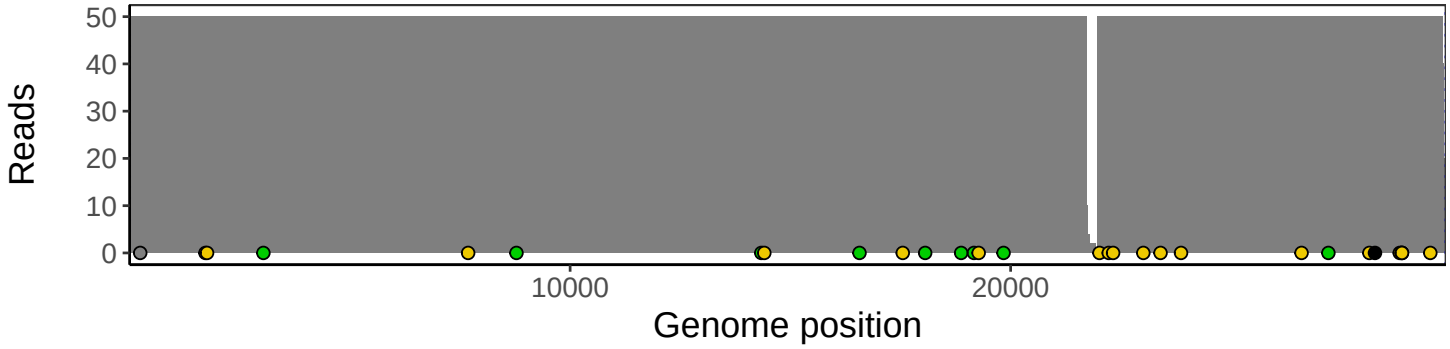


VSP0626-2 | 2021-01-22 | Saline | molpath-seq1 | genomes | single experiment

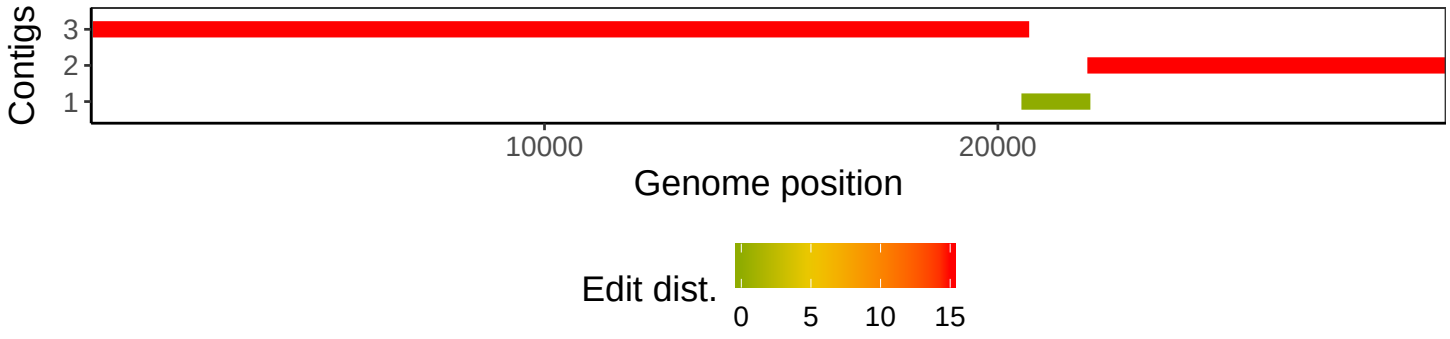
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

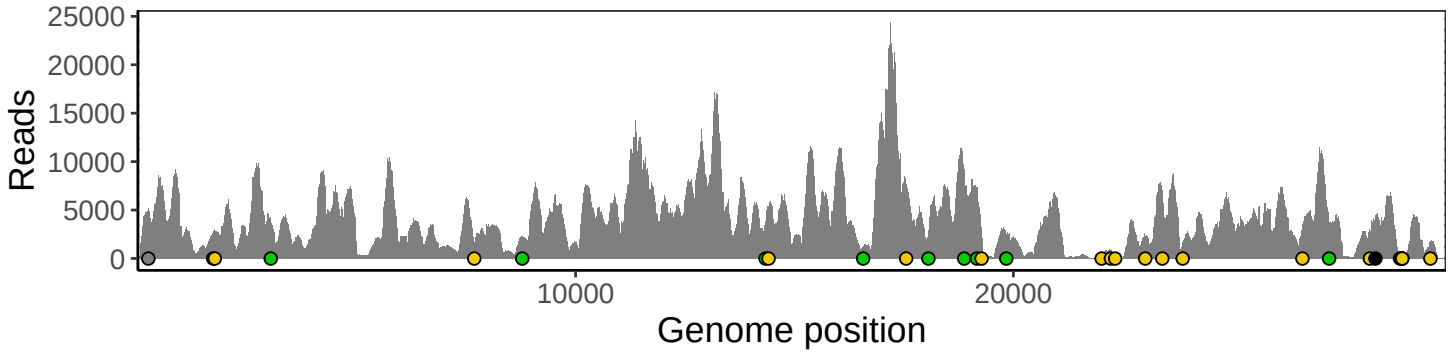


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

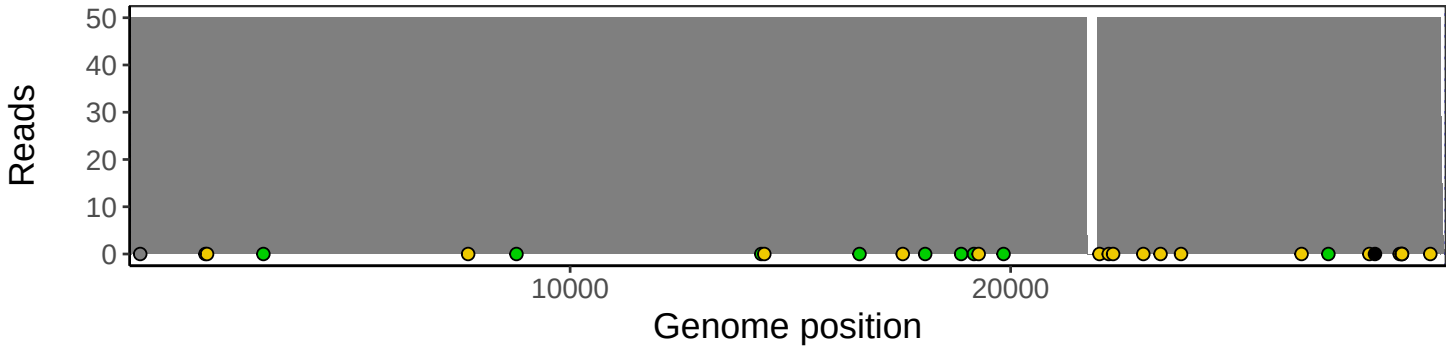


VSP0626-3 | 2021-01-22 | Saline | molpath-seq1 | genomes | single experiment

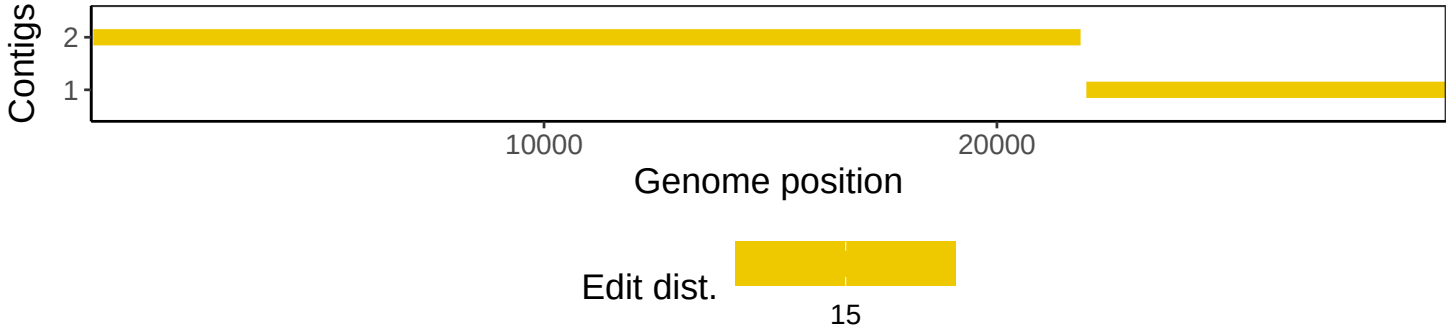
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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