COVID-19 subject MPCluster2-Seq7

2021-03-29

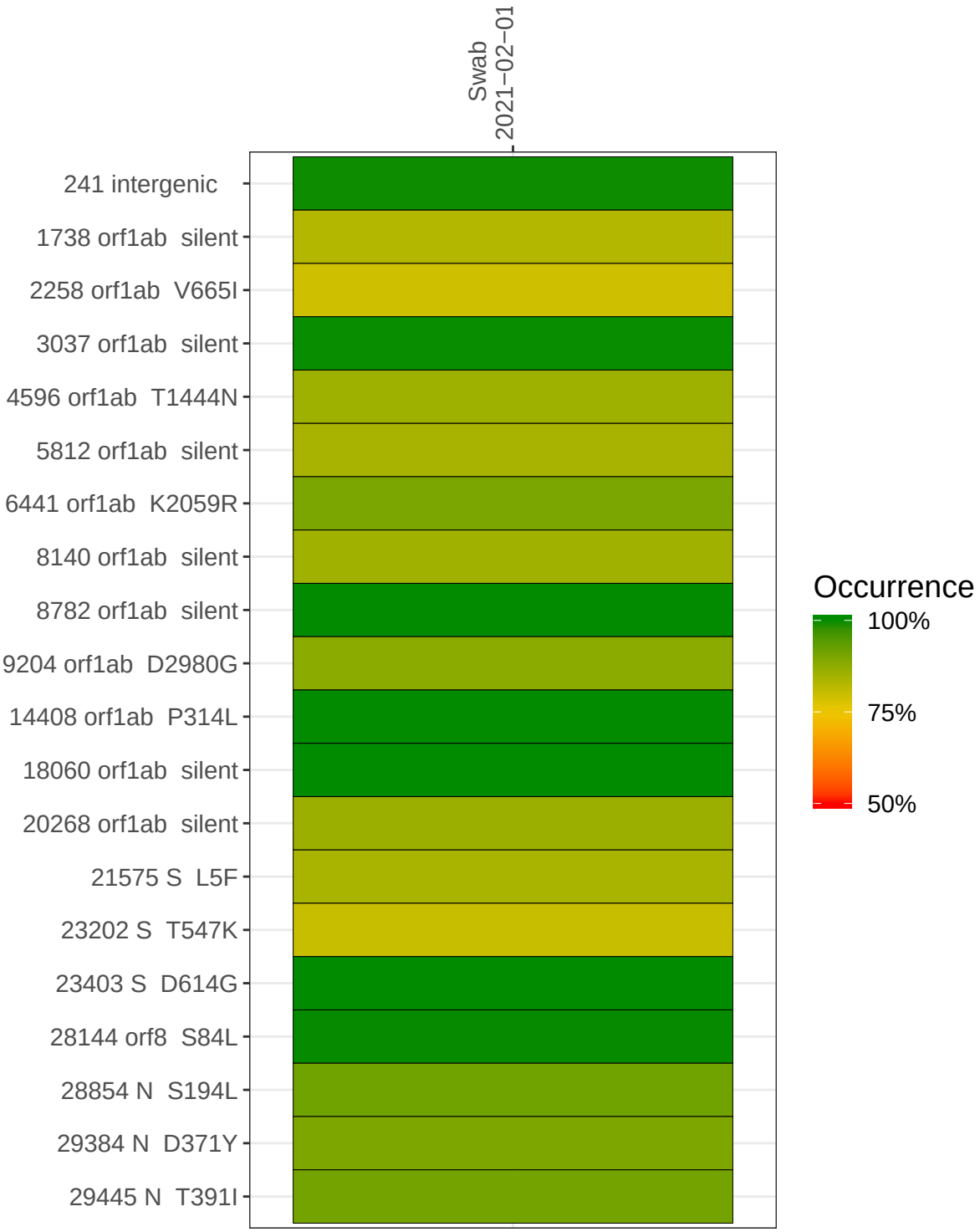
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

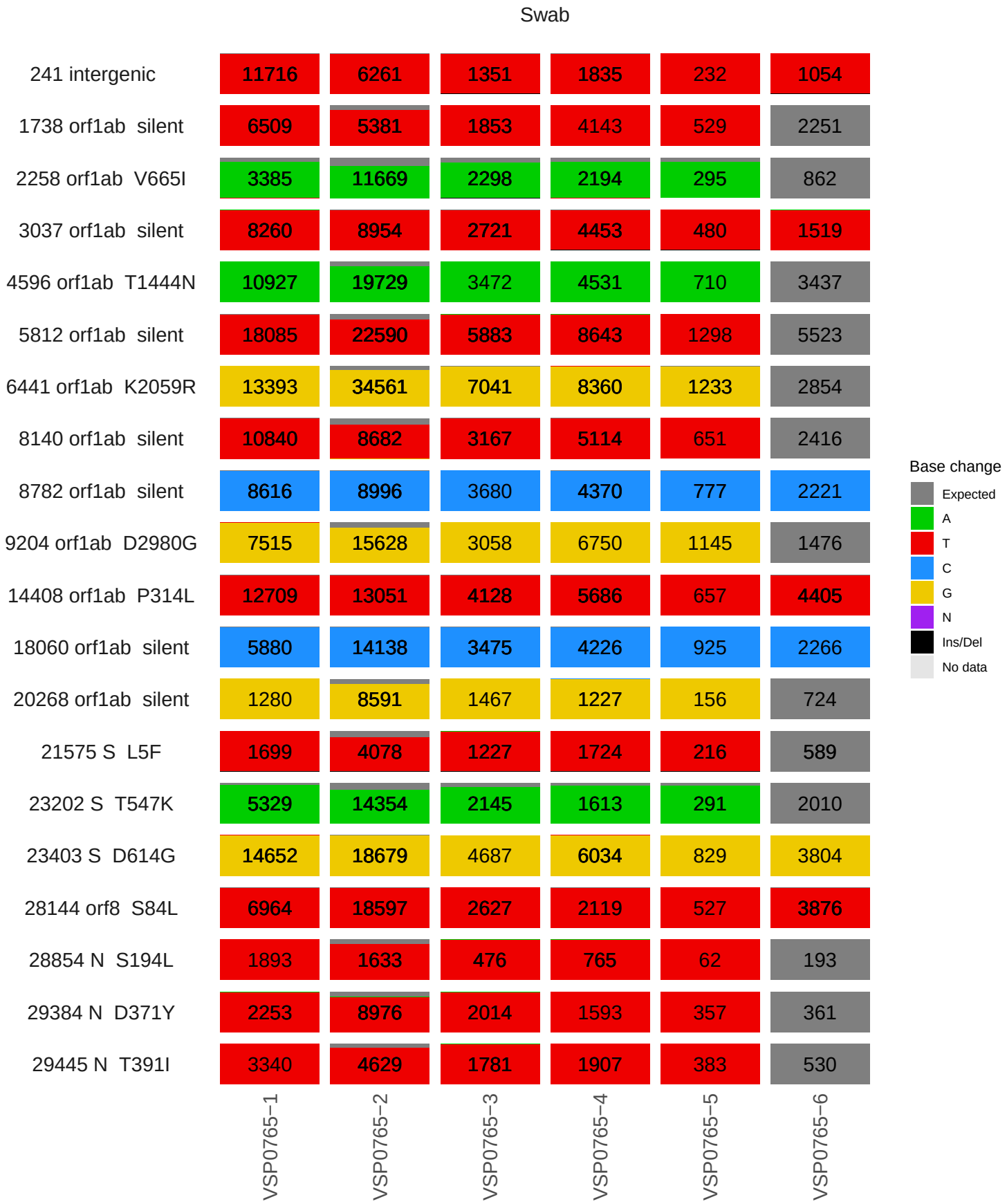
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0765	composite	NA	Swab	2021-02-01	29.96	B.1.234	99.9%	99.8%
VSP0765-1	single experiment	NA	Swab	2021-02-01	29.89	B.1.234	99.8%	99.8%
VSP0765-2	single experiment	NA	Swab	2021-02-01	29.88	B.1.234	99.9%	99.8%
VSP0765-3	single experiment	NA	Swab	2021-02-01	29.85	B.1.234	99.8%	99.8%
VSP0765-4	single experiment	NA	Swab	2021-02-01	29.84	B.1.234	99.8%	99.8%
VSP0765-5	single experiment	NA	Swab	2021-02-01	29.91	B.1.234	99.8%	99.6%
VSP0765-6	single experiment	NA	Swab	2021-02-01	29.43	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.

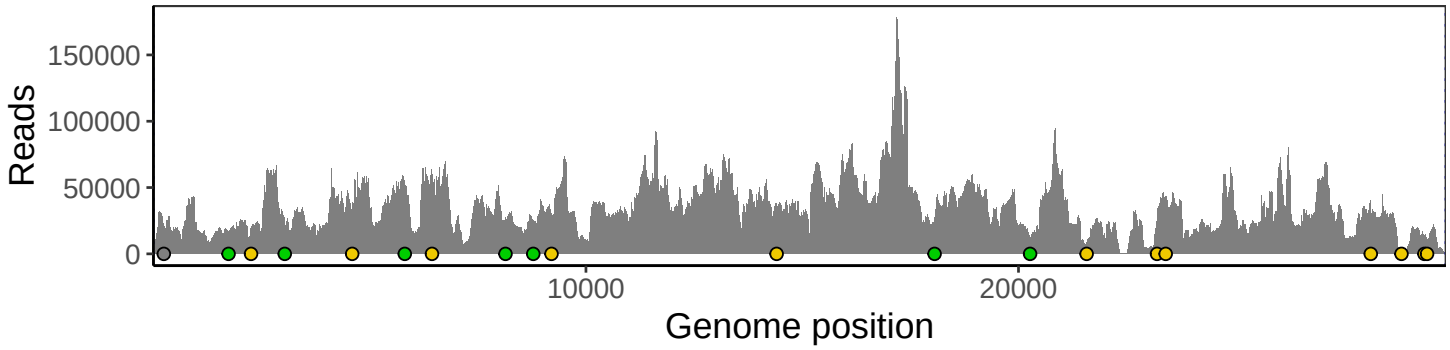



Swab
241 intergenic
1738 orf1ab silent
2258 orf1ab V665I
3037 orf1ab silent
4596 orf1ab T1444N
5812 orf1ab silent
6441 orf1ab K2059R
8140 orf1ab silent
8782 orf1ab silent
9204 orf1ab D2980G
14408 orf1ab P314L
18060 orf1ab silent
20268 orf1ab silent
21575 S L5F
23202 S T547K
23403 S D614G
28144 orf8 S84L
28854 N S194L
29384 N D371Y
29445 N T391I
11716 6261 1351 1835 232 1054
6509 5381 1853 4143 529 2251
3385 11669 2298 2194 295 862
8260 8954 2721 4453 480 1519
10927 19729 3472 4531 710 3437
18085 22590 5883 8643 1298 5523
13393 34561 7041 8360 1233 2854
10840 8682 3167 5114 651 2416
8616 8996 3680 4370 777 2221
7515 15628 3058 6750 1145 1476
12709 13051 4128 5686 657 4405
5880 14138 3475 4226 925 2266
1280 8591 1467 1227 156 724
1699 4078 1227 1724 216 589
5329 14354 2145 1613 291 2010
14652 18679 4687 6034 829 3804
6964 18597 2627 2119 527 3876
1893 1633 476 765 62 193
2253 8976 2014 1593 357 361
3340 4629 1781 1907 383 530
VSP0765−1
VSP0765−2
VSP0765−3
VSP0765−4
VSP0765−5
VSP0765−6
Base change
Expected
A
T
C
G
N
Ins/Del
No data

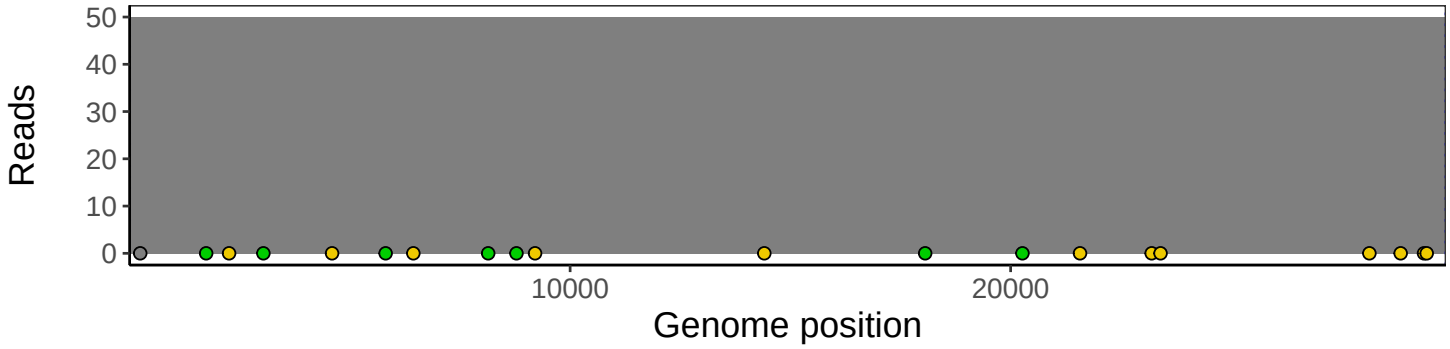
Analyses of individual experiments and composite results

VSP0765 | 2021-02-01 | Swab | MPCluster2-Seq7 | composite result

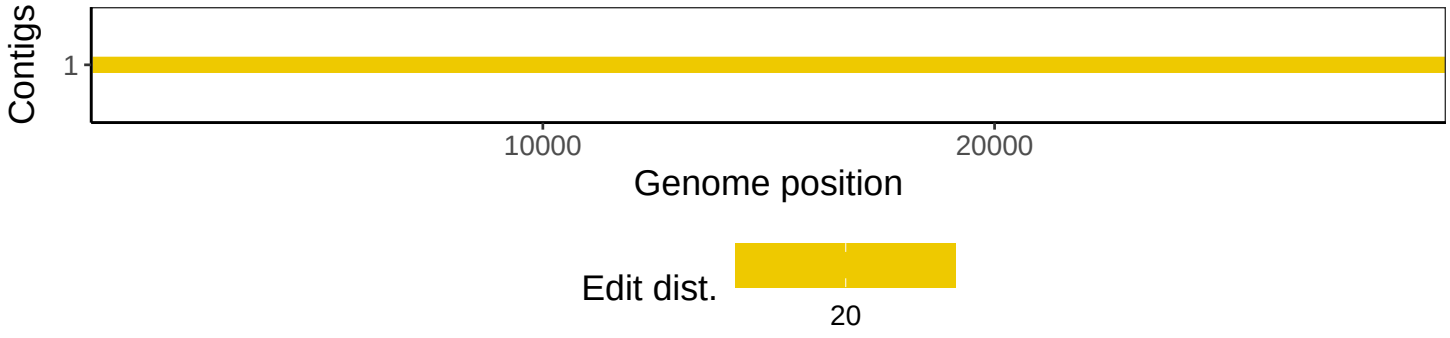
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

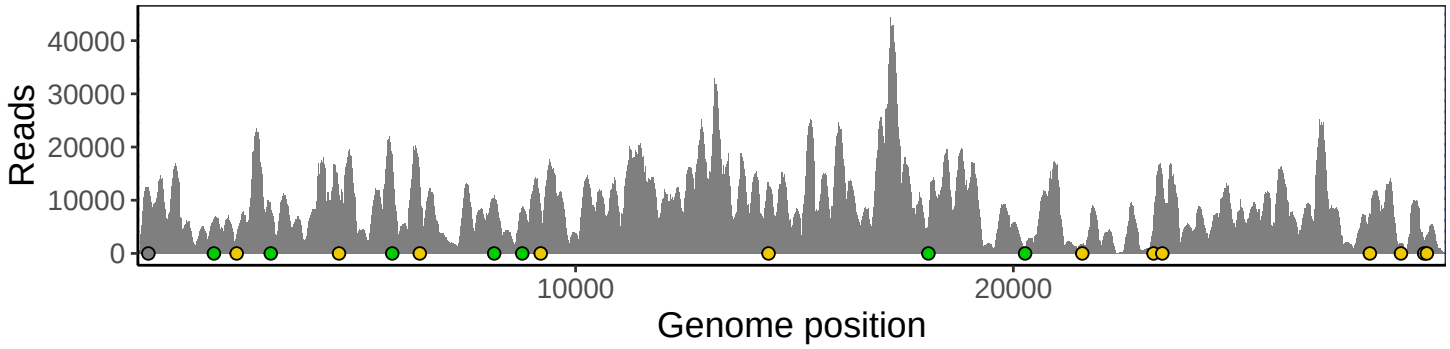


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

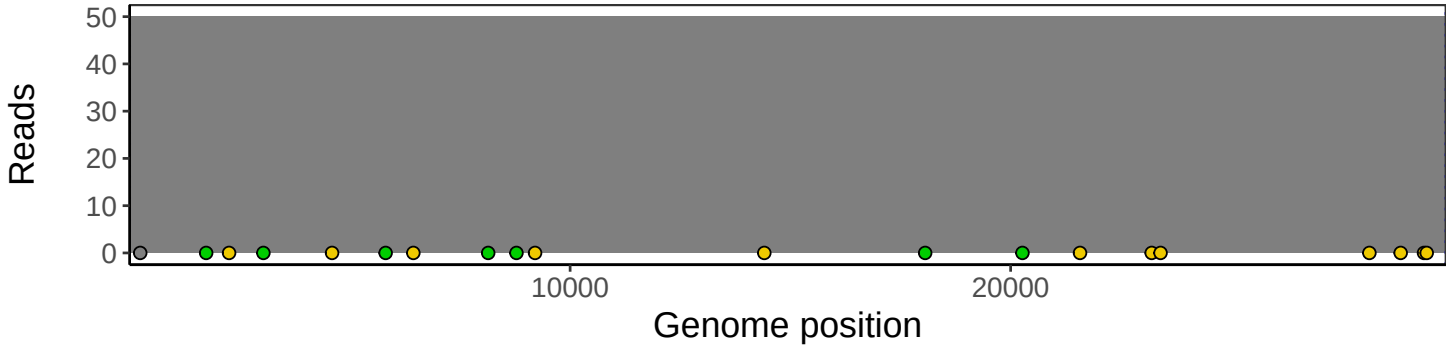


VSP0765-1 | 2021-02-01 | Swab | MPCluster2-Seq7 | genomes | single experiment

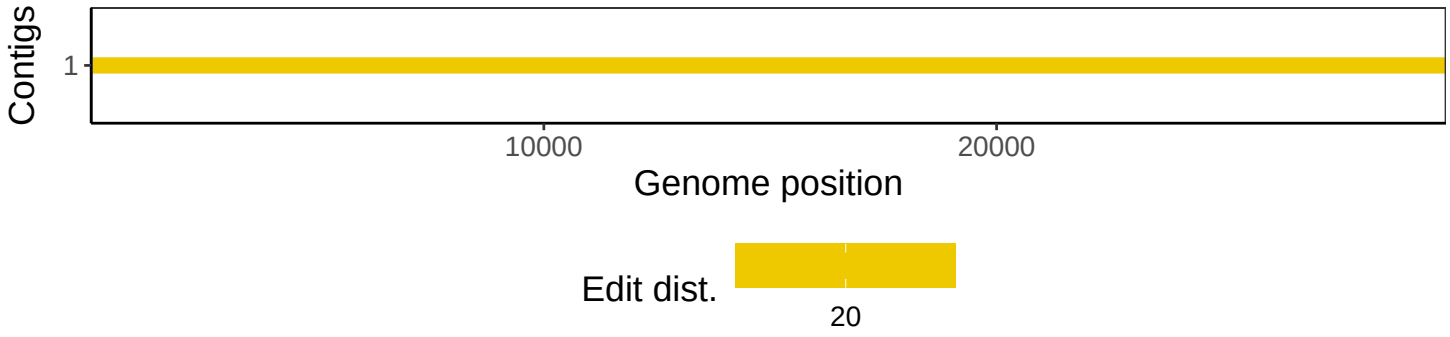
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

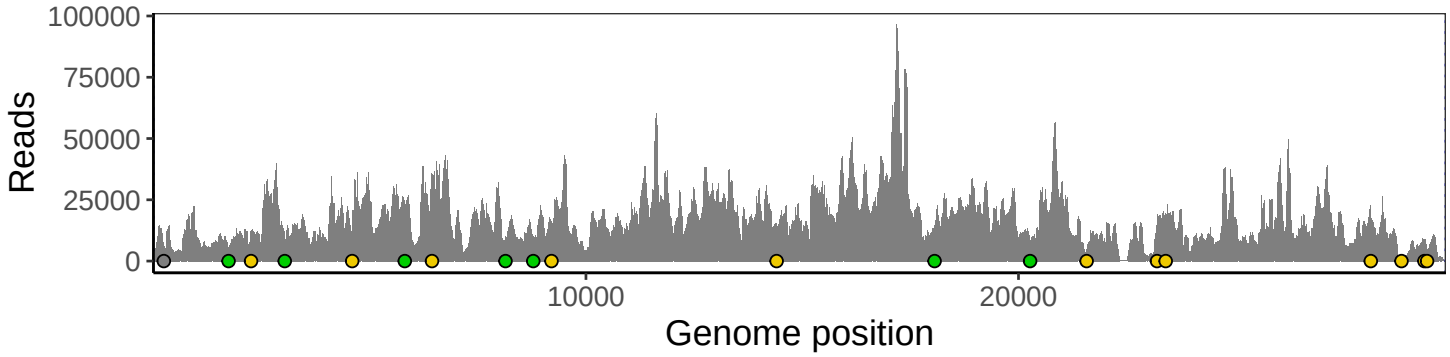


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

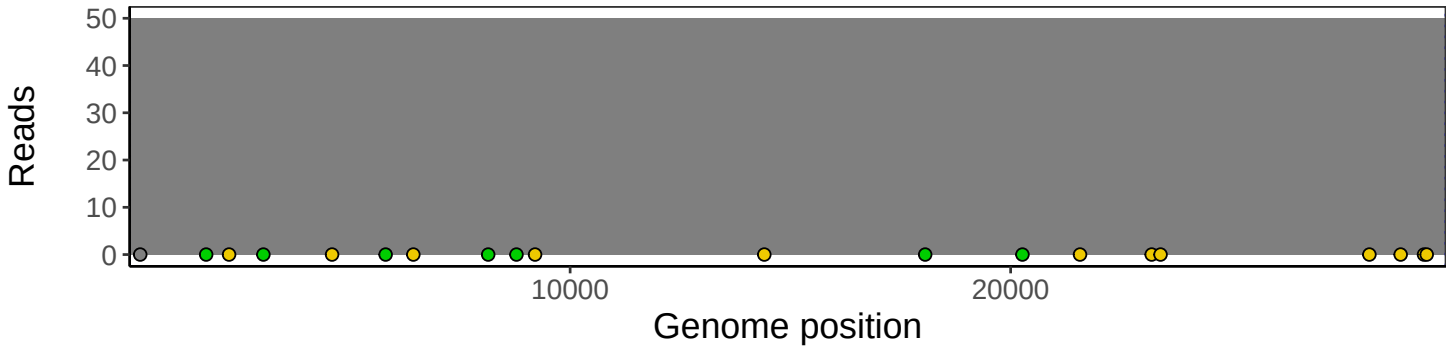


VSP0765-2 | 2021-02-01 | Swab | MPCluster2-Seq7 | genomes | single experiment

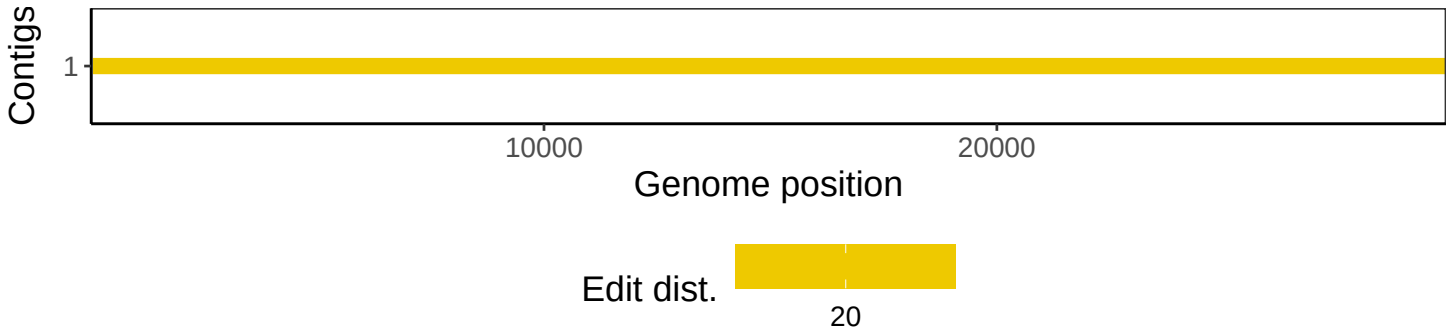
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

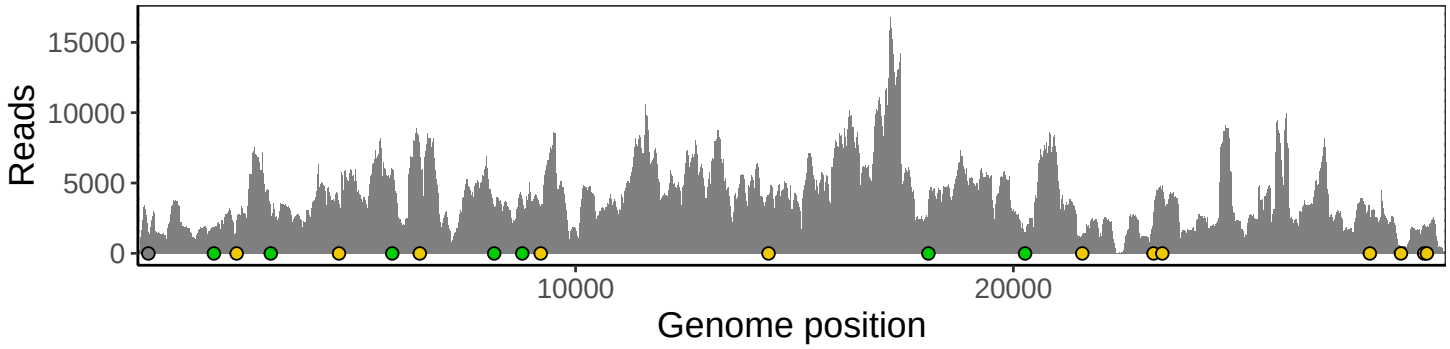


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

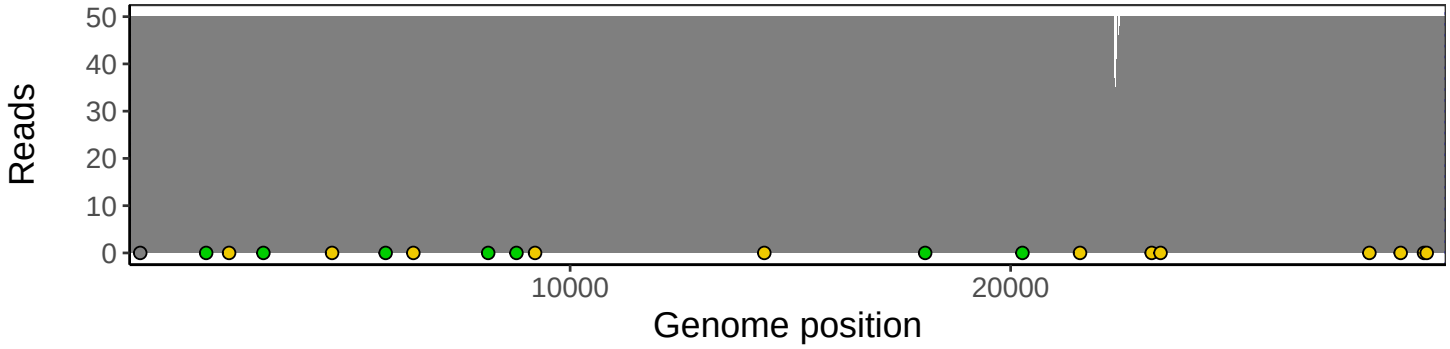


VSP0765-3 | 2021-02-01 | Swab | MPCluster2-Seq7 | genomes | single experiment

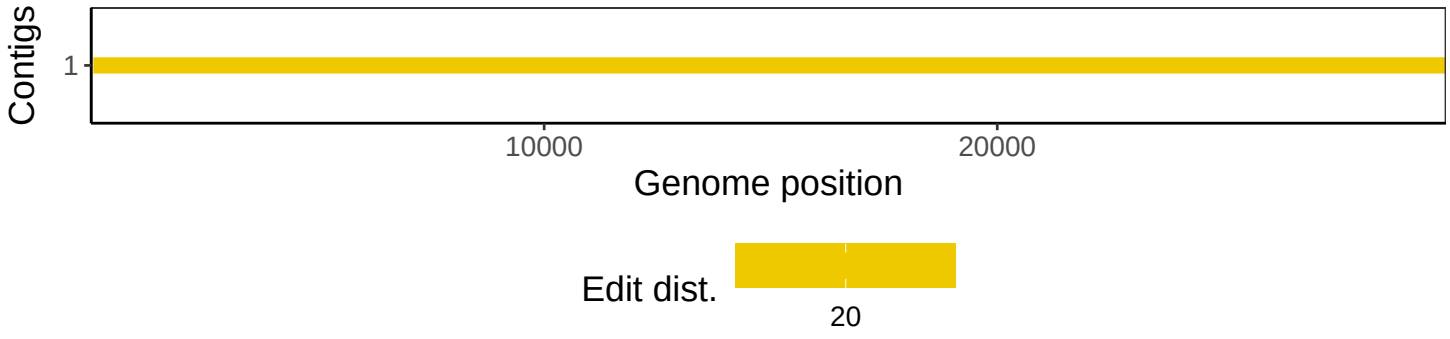
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

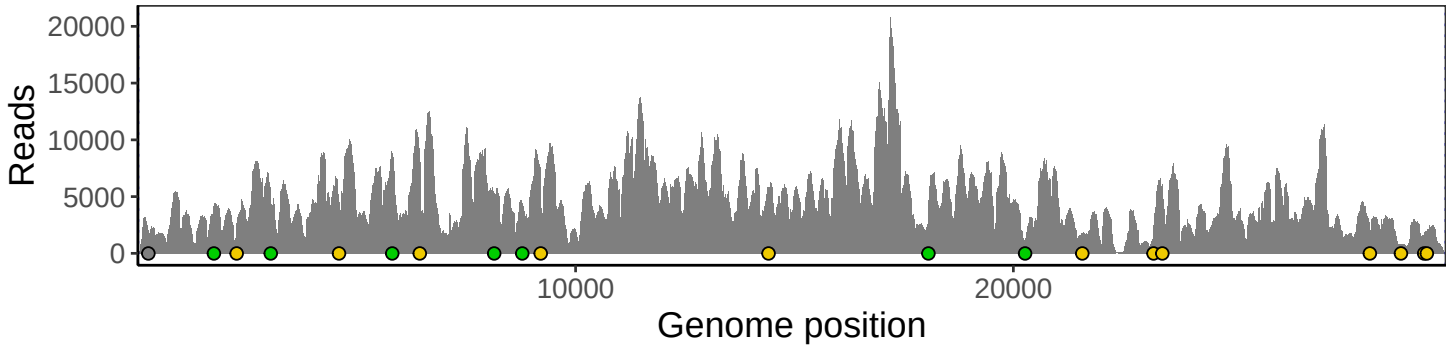


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

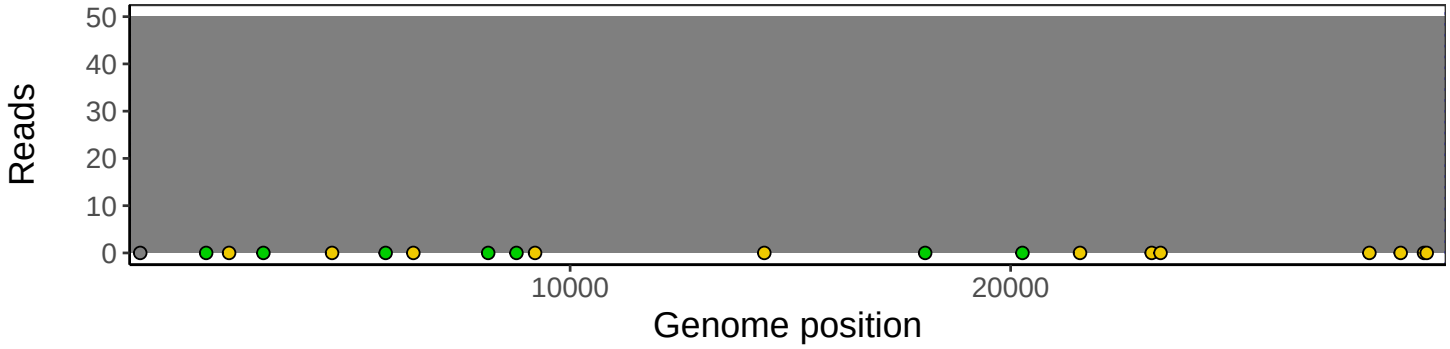


VSP0765-4 | 2021-02-01 | Swab | MPCluster2-Seq7 | genomes | single experiment

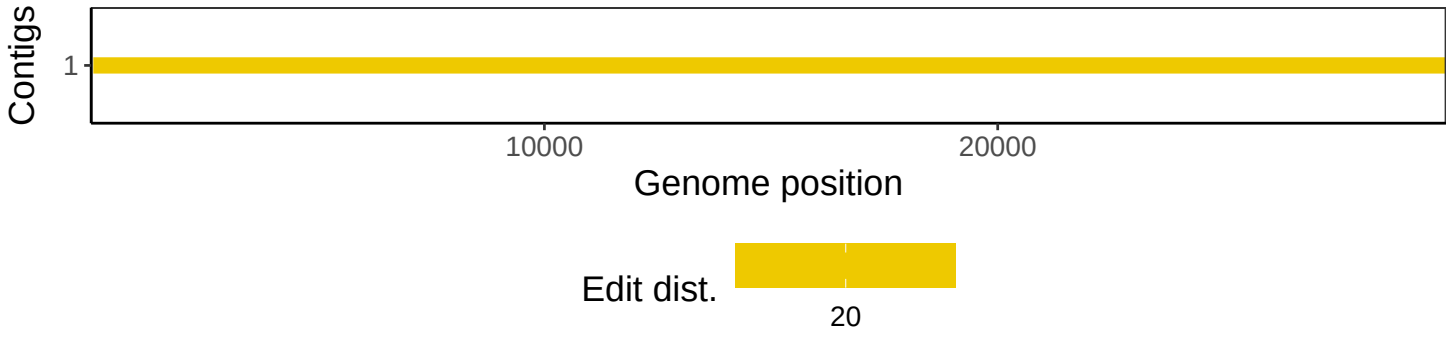
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

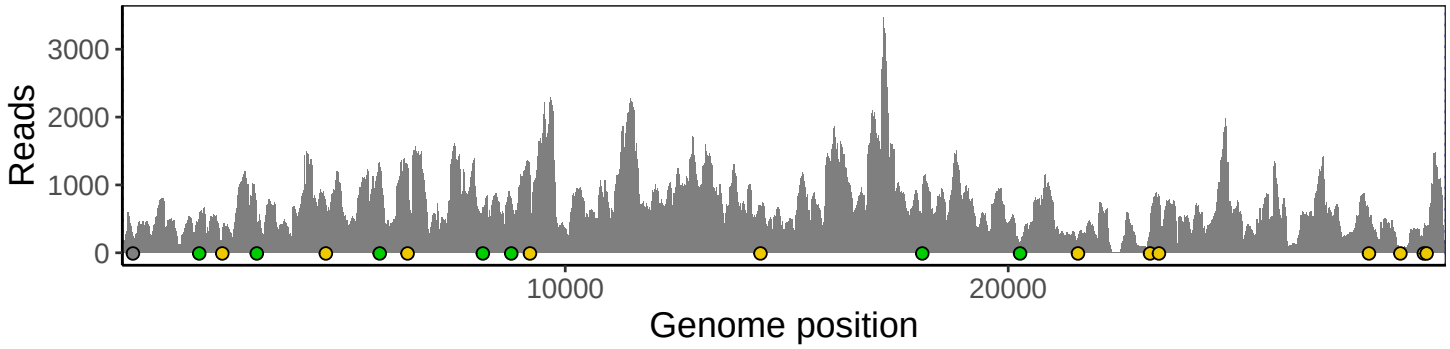


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

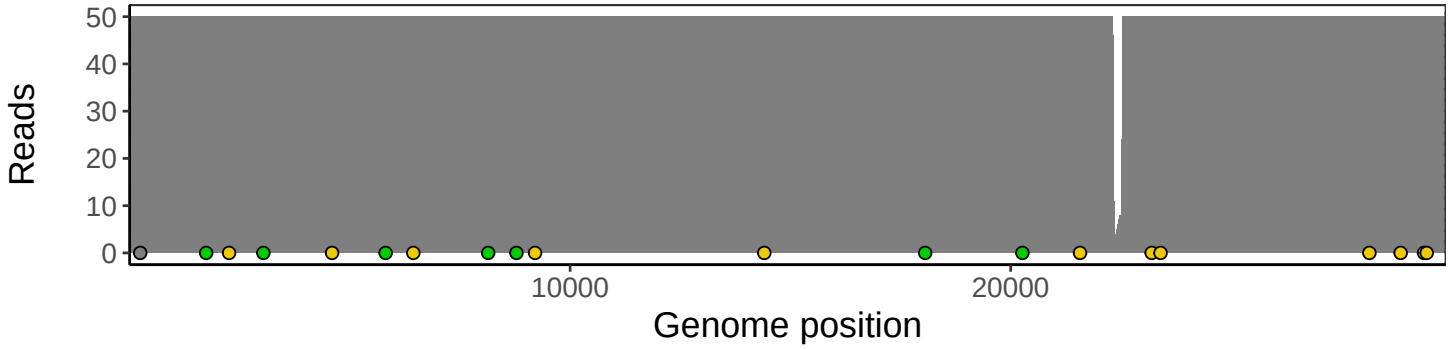


VSP0765-5 | 2021-02-01 | Swab | MPCluster2-Seq7 | genomes | single experiment

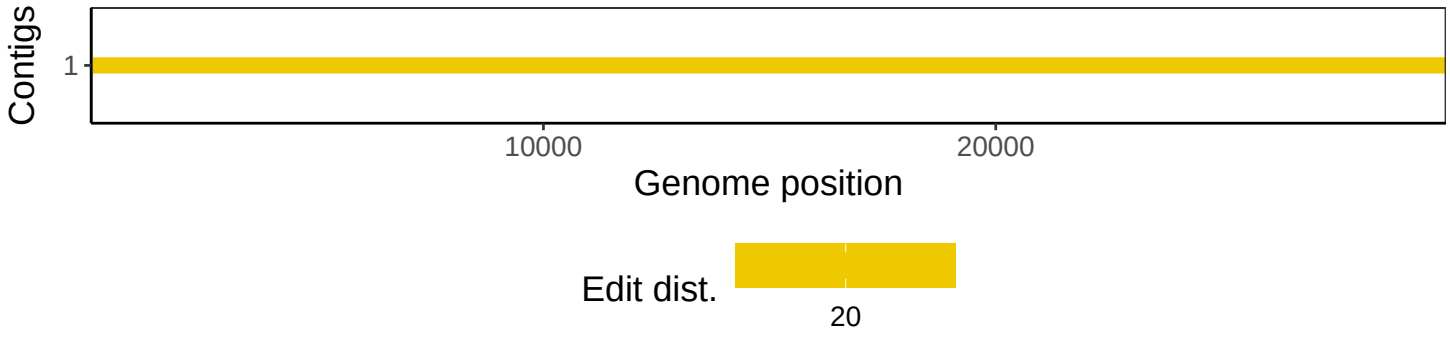
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

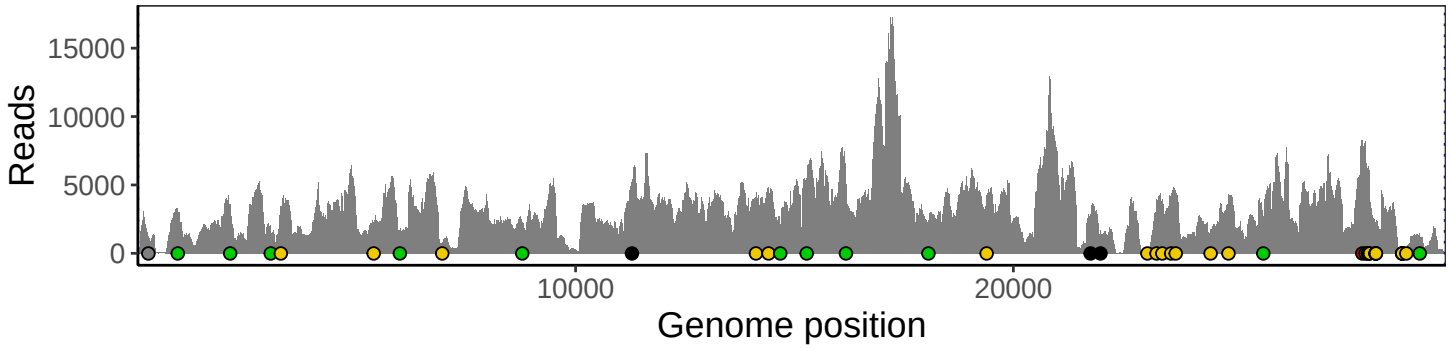


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

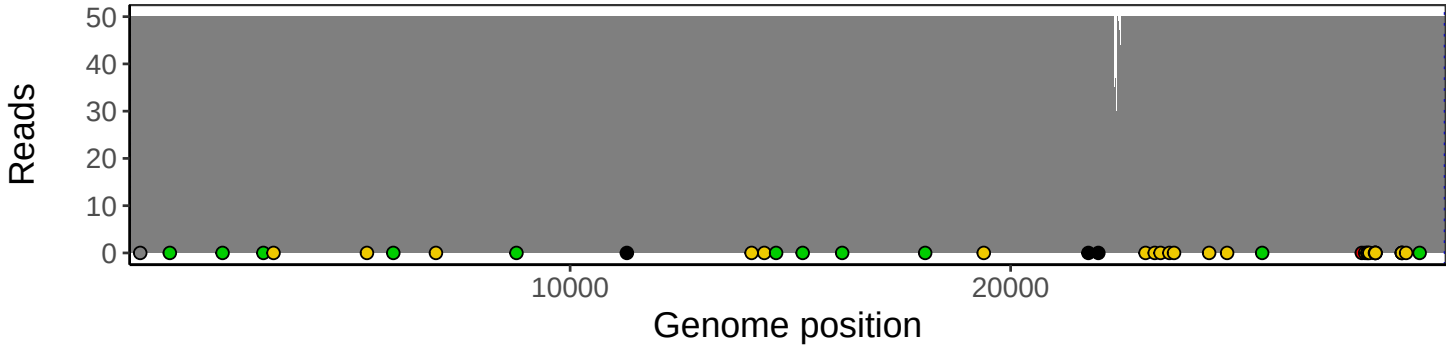


VSP0765-6 | 2021-02-01 | Swab | MPCluster2-Seq7 | genomes | single experiment

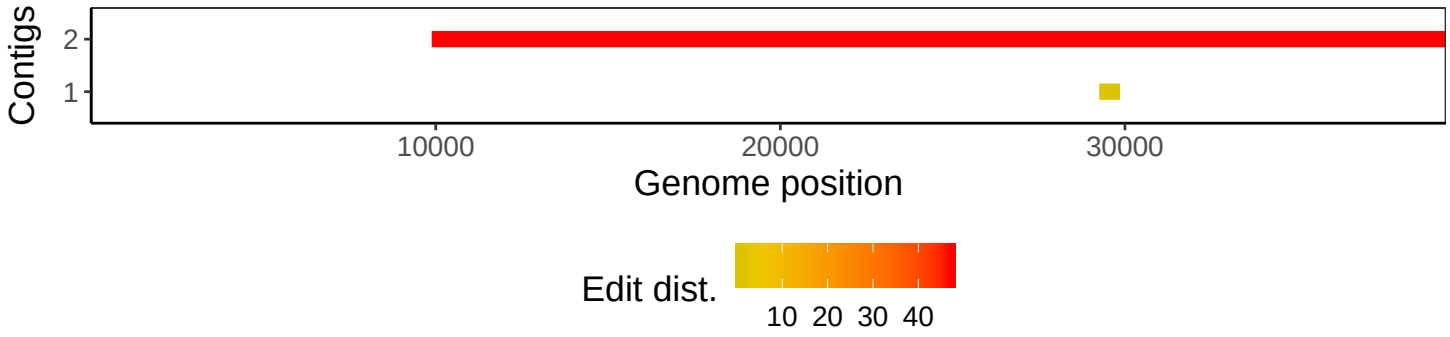
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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