

# COVID-19 subject 228

2021-06-23

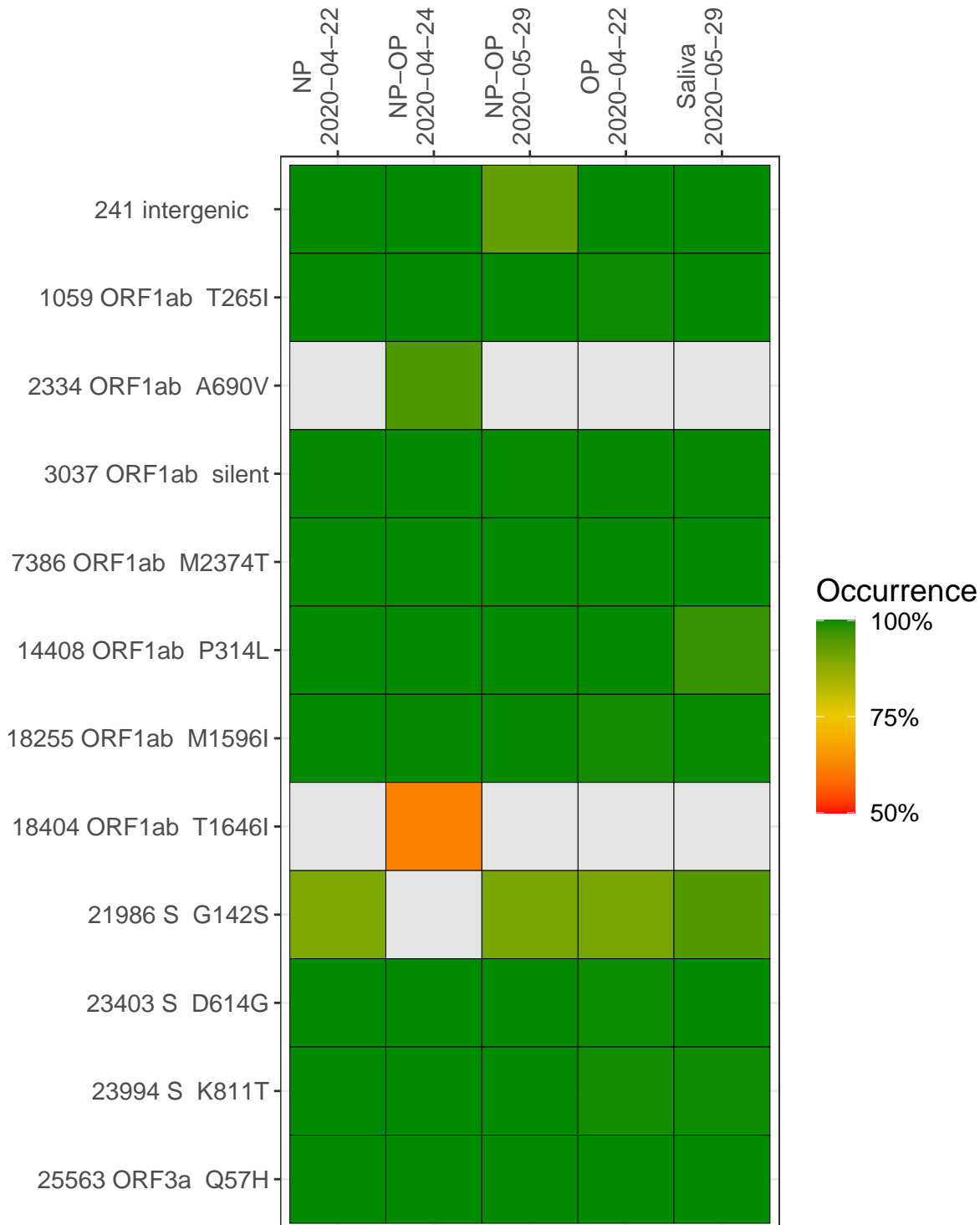
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP0021	composite	NA	NP	2020-04-22	29.82	B.1	99.7%	99.7%
VSP0022	composite	NA	OP	2020-04-22	29.68	B.1	99.7%	99.6%
VSP0069	composite	NA	NP-OP	2020-04-24	18.98	B.1	98.5%	98.5%
VSP0187	composite	NA	ETA	2020-05-29	0.30	NA	14.3%	0.3%
VSP0188	composite	NA	NP-OP	2020-05-29	29.90	B.1	99.8%	99.8%
VSP0021-1m	single experiment	NA	NP	2020-04-22	29.82	B.1	99.7%	99.7%
VSP0021-2	single experiment	1.10e+04	NP	2020-04-22	29.82	B.1	99.7%	99.7%
VSP0022-1a	single experiment	2.68e+05	OP	2020-04-22	1.80	NA	81.4%	58.6%
VSP0022-2	single experiment	1.34e+06	OP	2020-04-22	0.24	NA	6.2%	0.0%
VSP0022-3	single experiment	1.34e+06	OP	2020-04-22	29.68	B.1	99.7%	99.6%
VSP0069-1	single experiment	9.75e+01	NP-OP	2020-04-24	6.54	NA	92.3%	88.7%
VSP0069-2	single experiment	9.75e+01	NP-OP	2020-04-24	9.23	NA	90.6%	90.2%
VSP0069-3	single experiment	9.75e+01	NP-OP	2020-04-24	6.37	NA	87.2%	85.9%
VSP0187-1	single experiment	6.02e+01	ETA	2020-05-29	0.30	NA	12.3%	0.3%
VSP0187-2	single experiment	3.01e+02	ETA	2020-05-29	NA	NA	2.4%	0.1%
VSP0188-1	single experiment	2.04e+03	NP-OP	2020-05-29	9.60	NA	93.7%	93.6%
VSP0188-2	single experiment	1.02e+04	NP-OP	2020-05-29	22.61	B.1	99.2%	99.2%
VSP0188-3	single experiment	1.02e+04	NP-OP	2020-05-29	29.87	B.1	99.8%	99.8%
VSP0189-1	single experiment	8.51e+04	Saliva	2020-05-29	29.82	B.1	99.9%	99.8%

## Variants shared across samples

The heat map below shows how variants (reference genome /home/common/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.



	NP 2020-04-22		NP-OP 2020-04-24			NP-OP 2020-05-29			OP 2020-04-22			Saliva 20-05-
241 intergenic	1318	6637	3767	2837	0	1763	4712	4037	13	0	791	984
1059 ORF1ab T265I	977	2617	8	1097	7774	2023	1862	2417	3	0	296	561
2334 ORF1ab A690V	1532	2603	470	0	9558	0	1134	544	2	0	298	560
3037 ORF1ab silent	1406	4820	177	1505	0	1530	2146	5459	0	0	1019	1354
7386 ORF1ab M2374T	741	2370	4711	2613	3182	2724	330	1029	8	0	496	580
14408 ORF1ab P314L	2107	10066	2768	3936	13819	2413	6673	6092	8	0	321	2634
18255 ORF1ab M1596I	1123	7370	1849	1779	8612	2857	2067	7506	0	0	384	1101
18404 ORF1ab T1646I	816	10114	4050	258	8295	1900	1890	2706	14	0	458	565
21986 S G142S	531	1102	3623	1453	1601	416	722	945	0	0	103	470
23403 S D614G	1253	8797	9973	14388	6	11269	5750	22032	69	0	956	2655
23994 S K811T	1326	3891	0	2432	5439	2779	2217	1651	19	2	577	508
25563 ORF3a Q57H	865	6637	2	8274	16277	3332	4022	5763	8	0	606	1848
	VSP0021-1m	VSP0021-2	VSP0069-1	VSP0069-2	VSP0069-3	VSP0188-1	VSP0188-2	VSP0188-3	VSP0022-1a	VSP0022-2	VSP0022-3	VSP0189-1

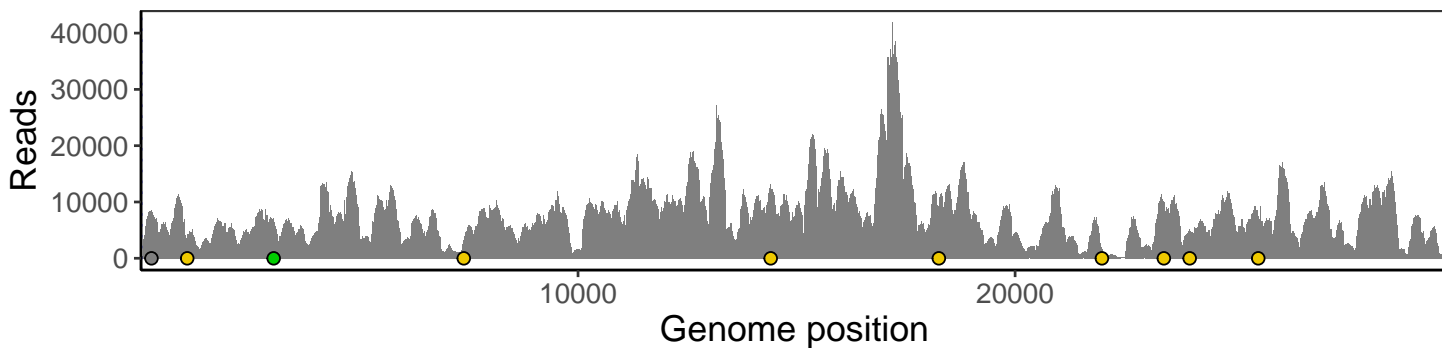
Base change



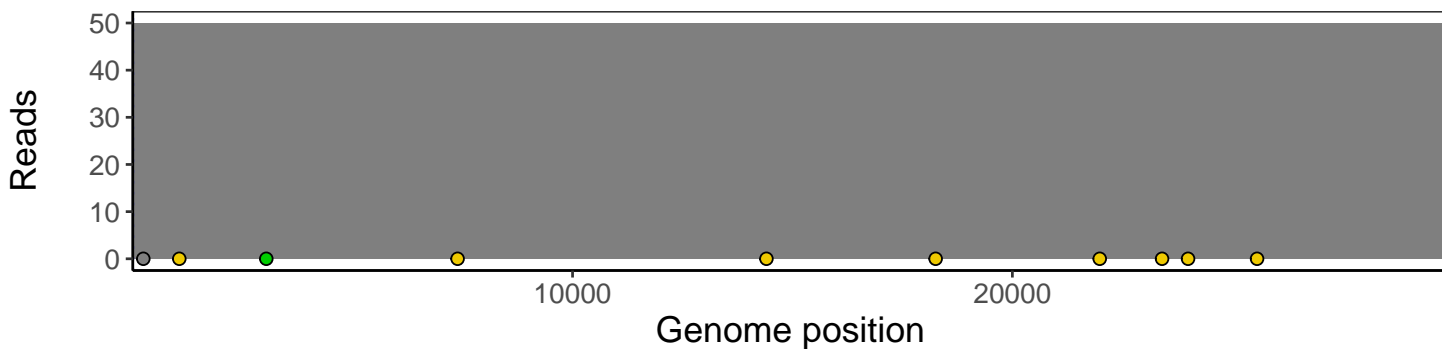
## Analyses of individual experiments and composite results

VSP0021 | 2020-04-22 | NP | 228n | composite result

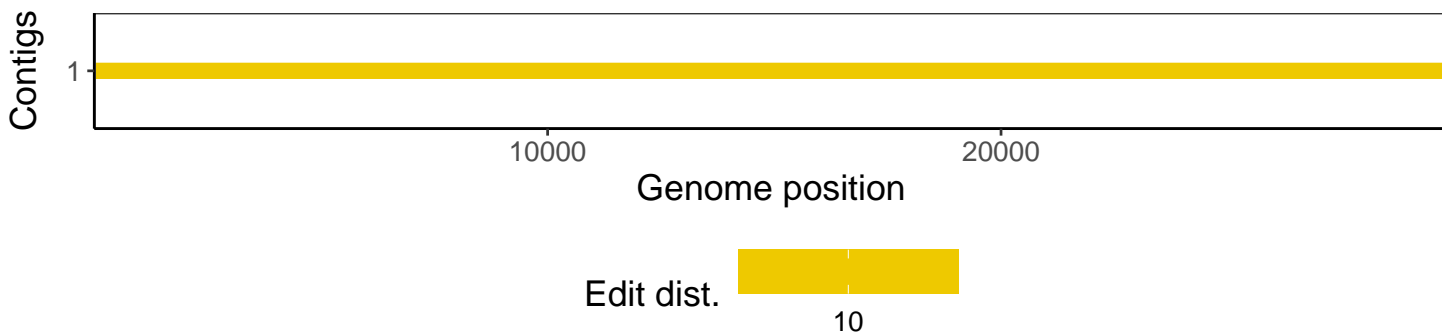
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according to variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



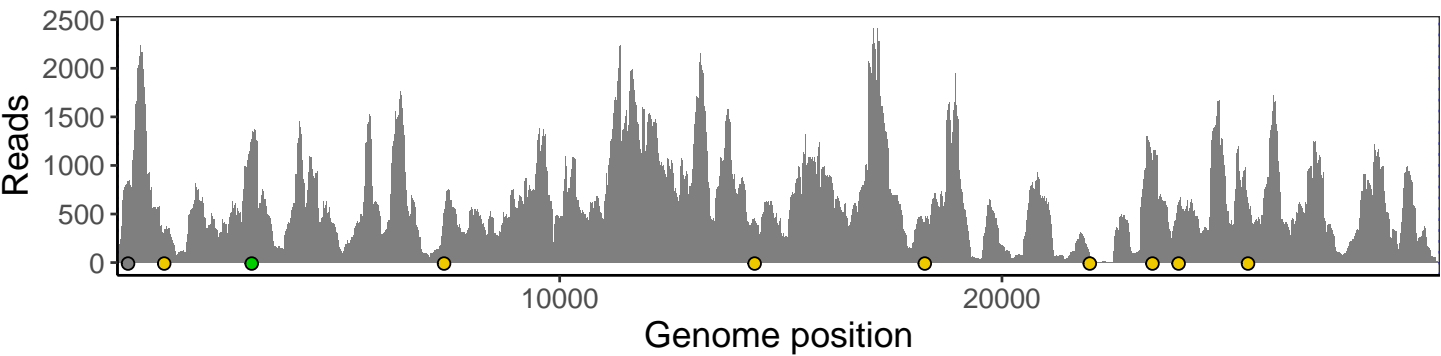
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



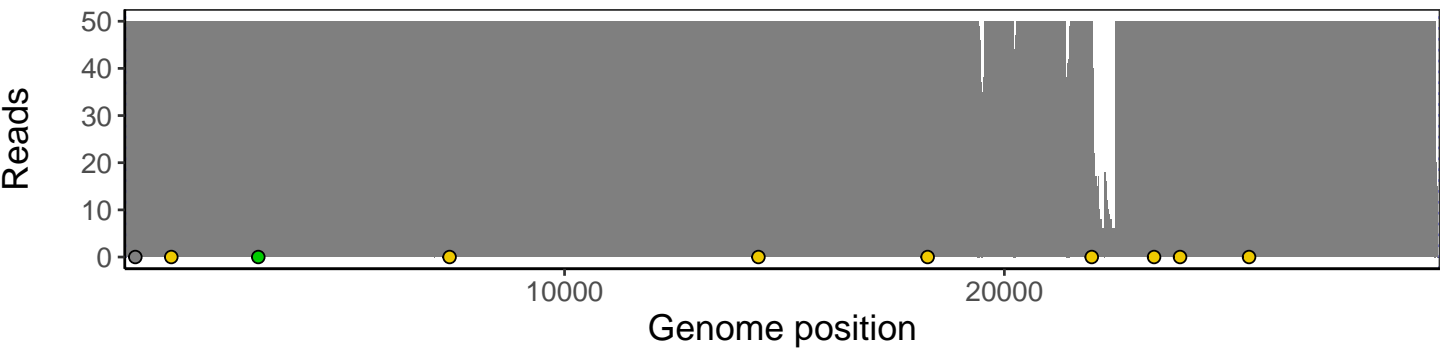
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



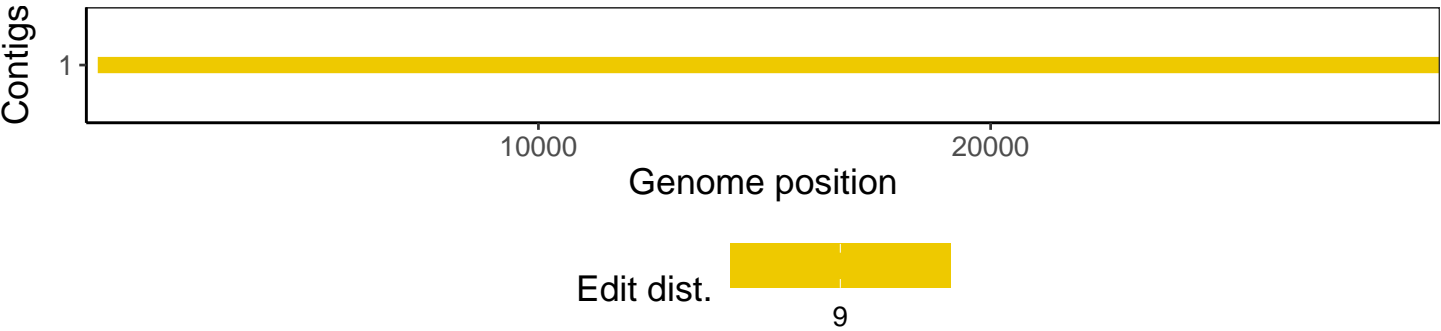
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



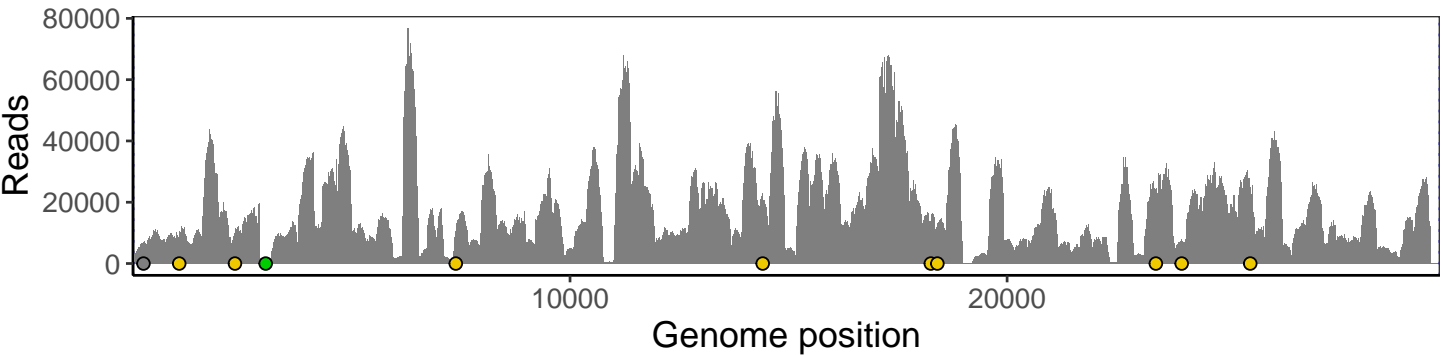
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



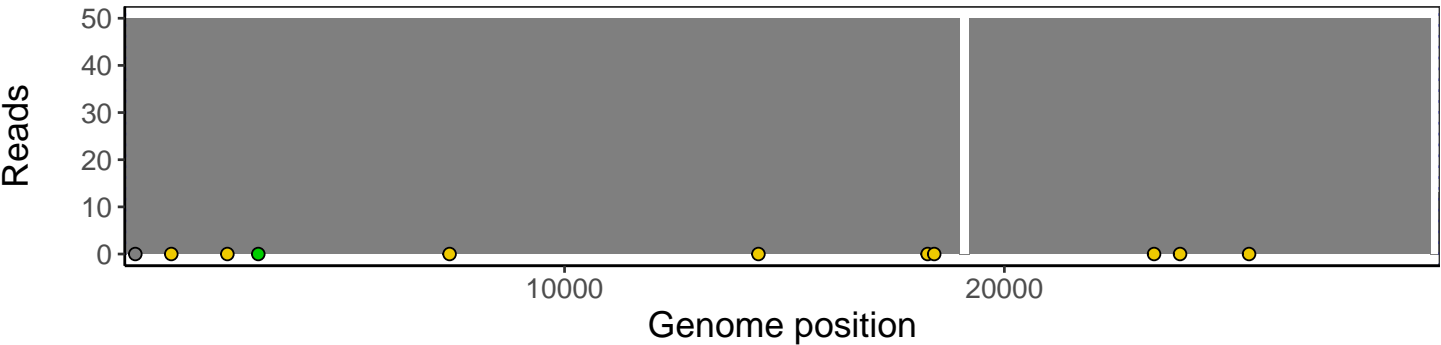
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



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.



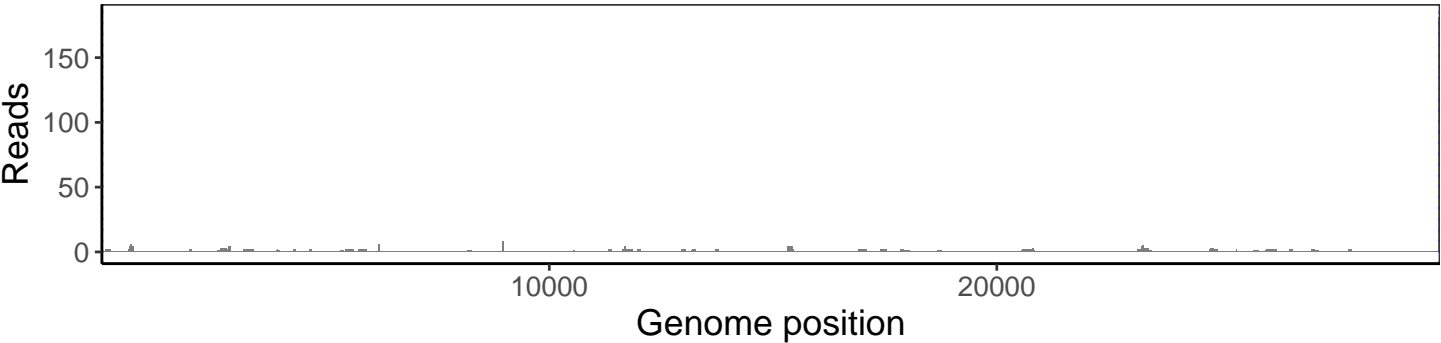
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



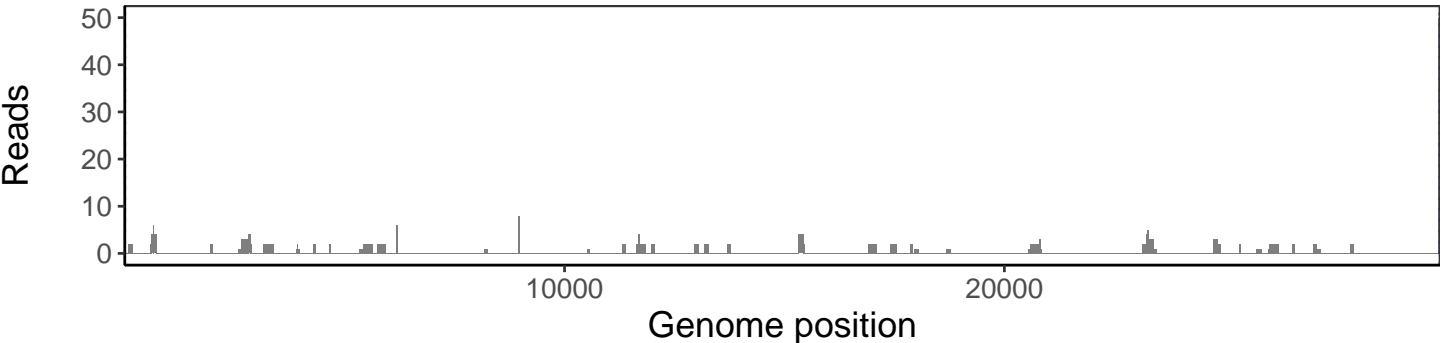
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



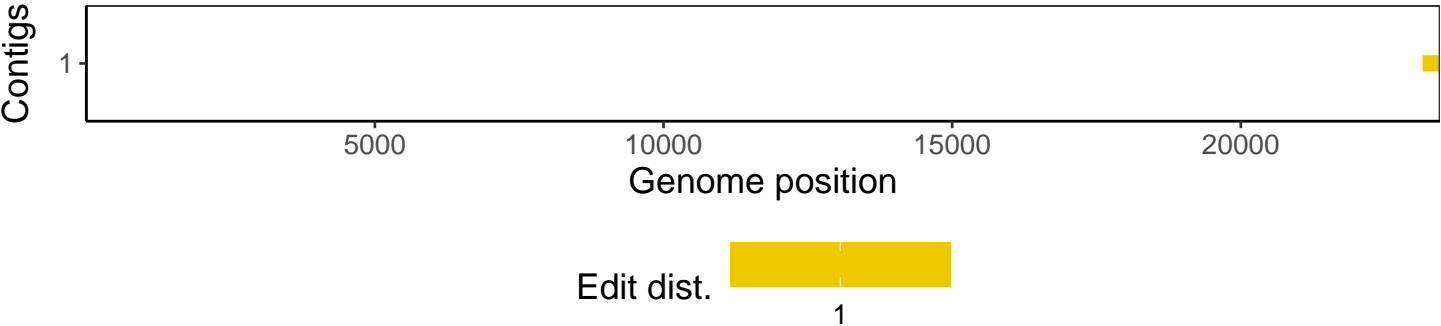
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



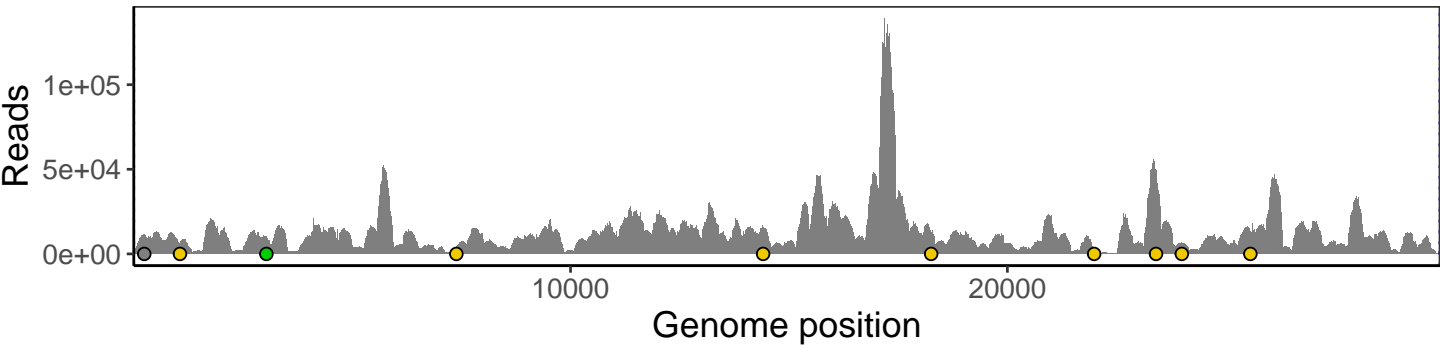
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



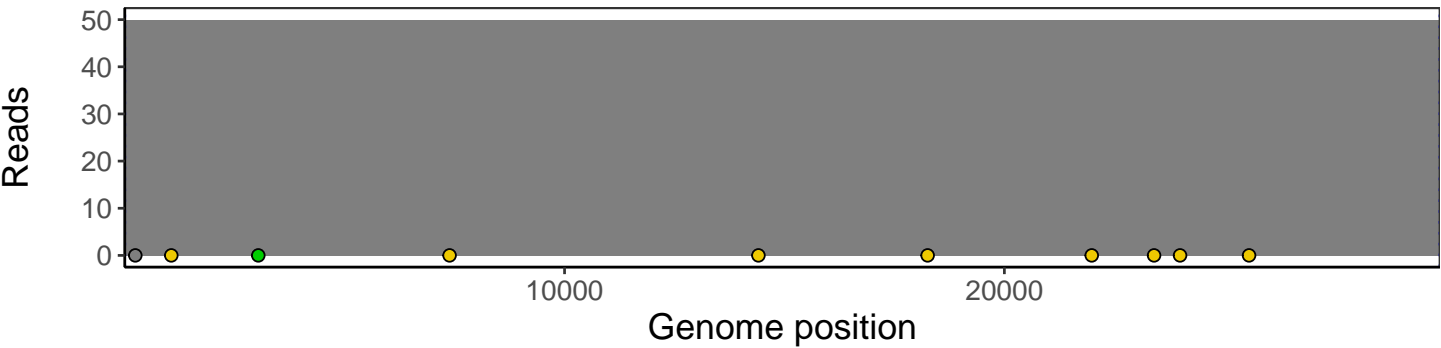
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



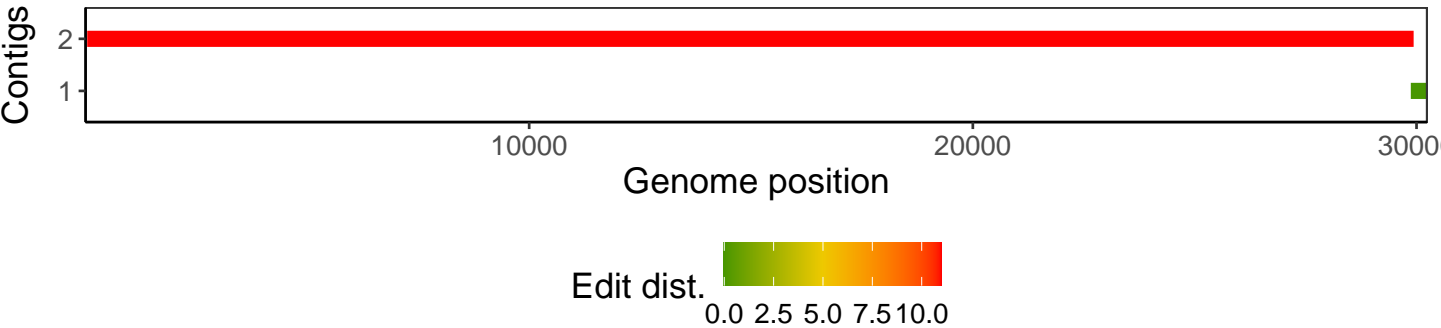
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

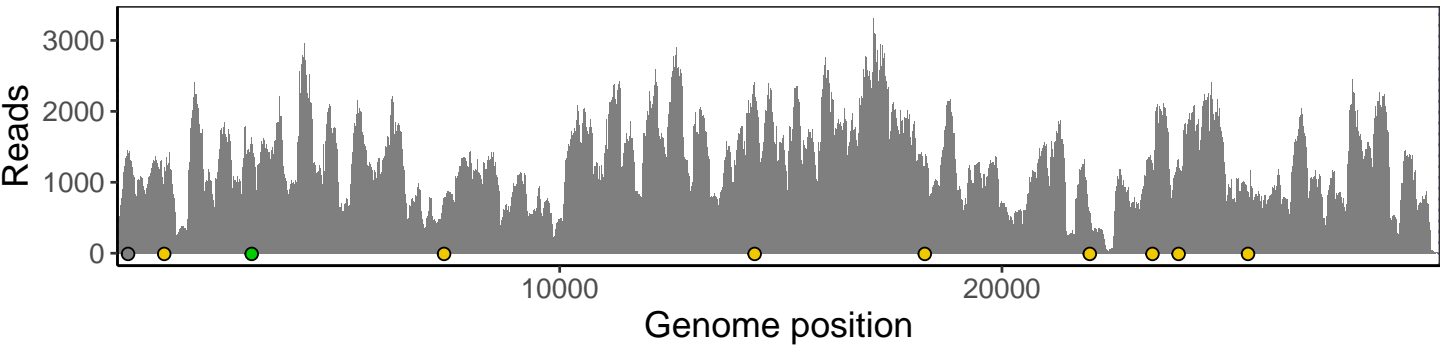


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

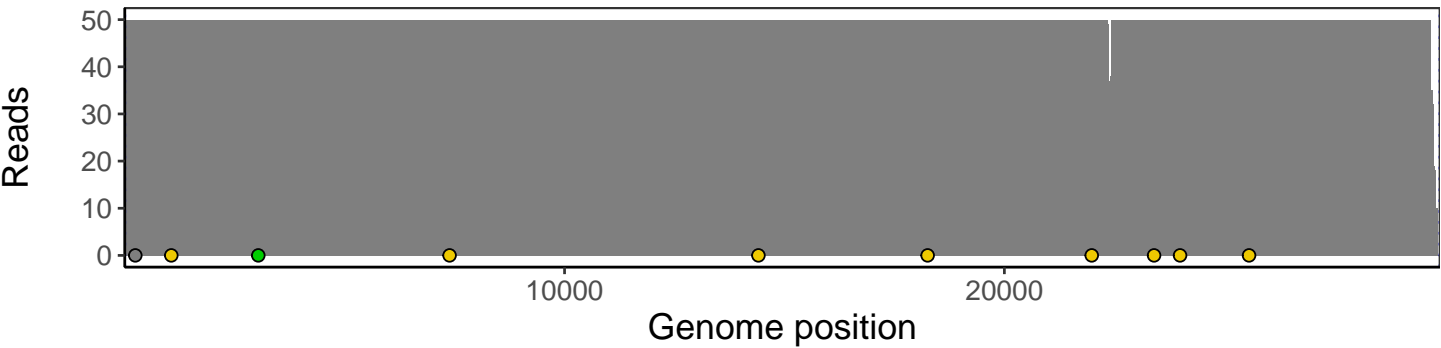




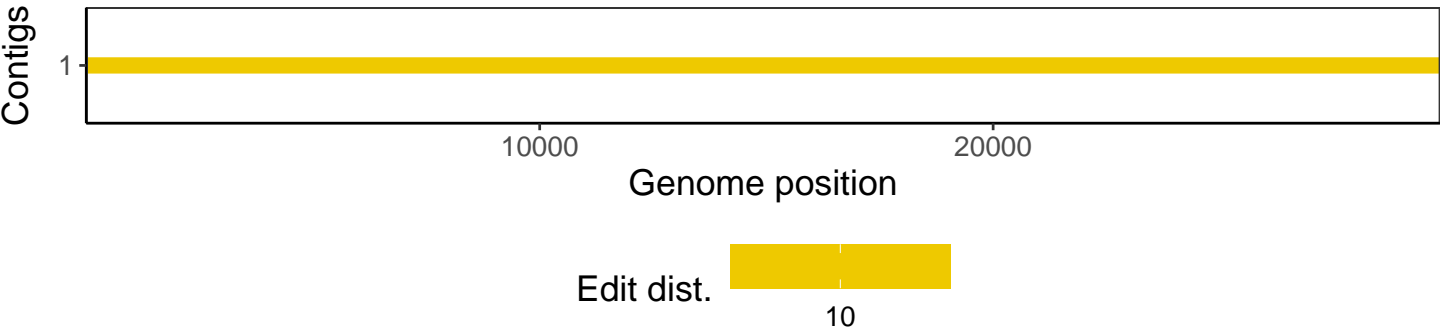
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



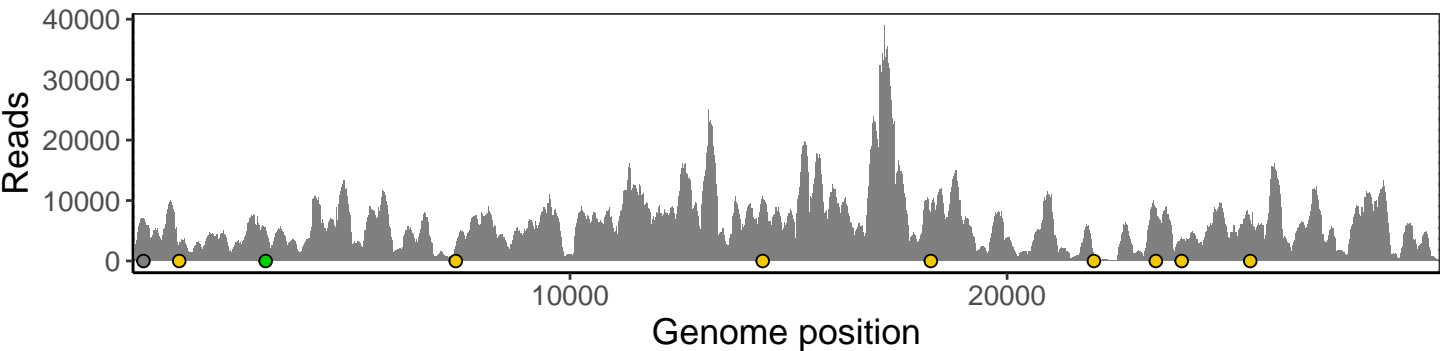
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



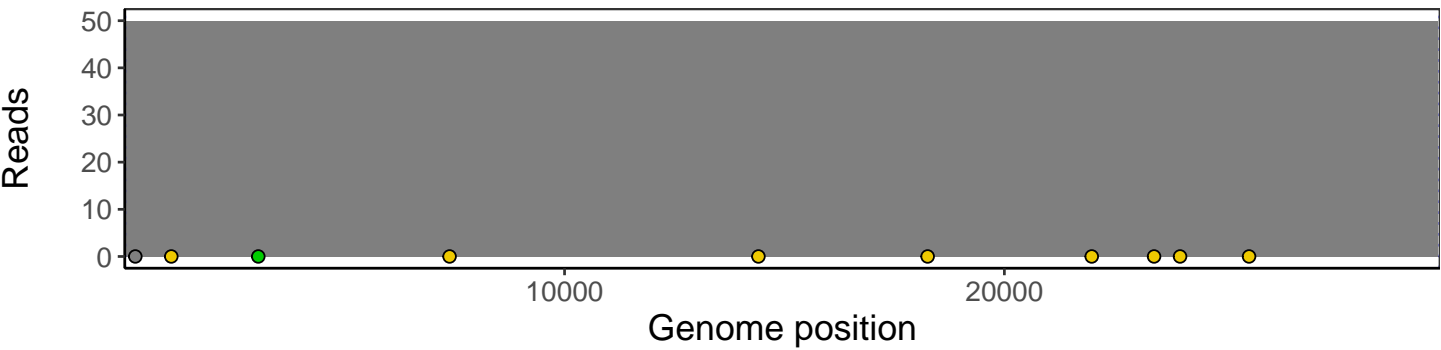
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



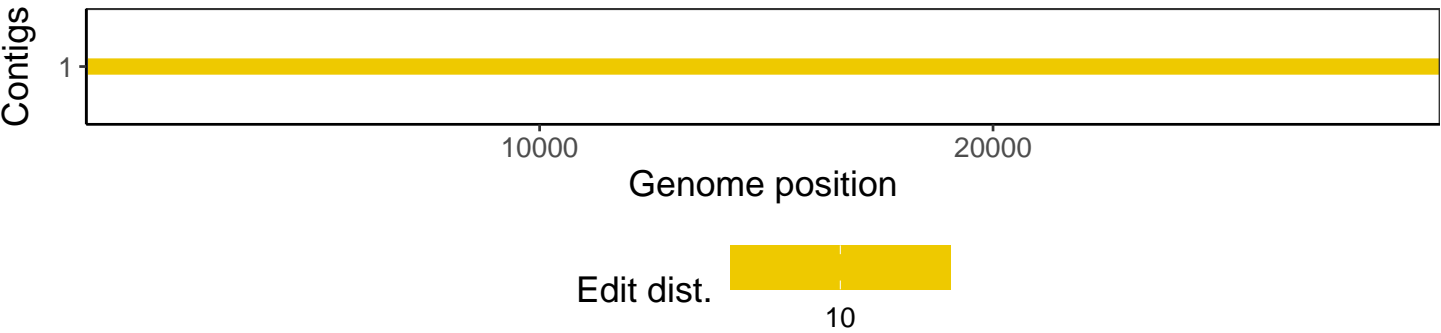
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



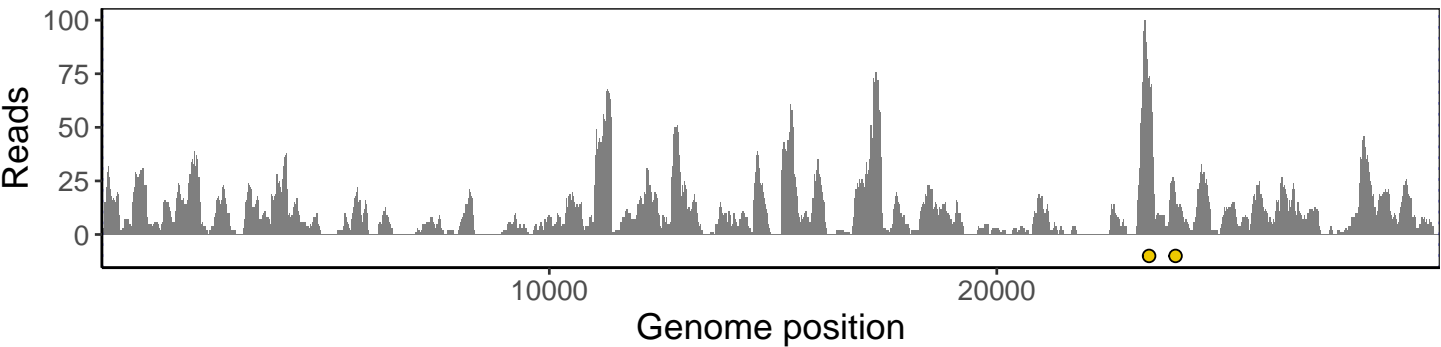
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



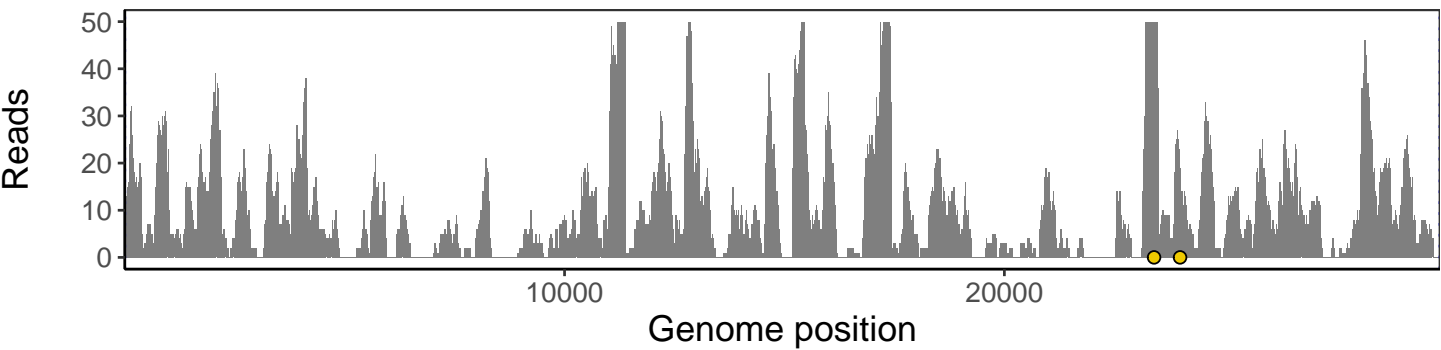
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



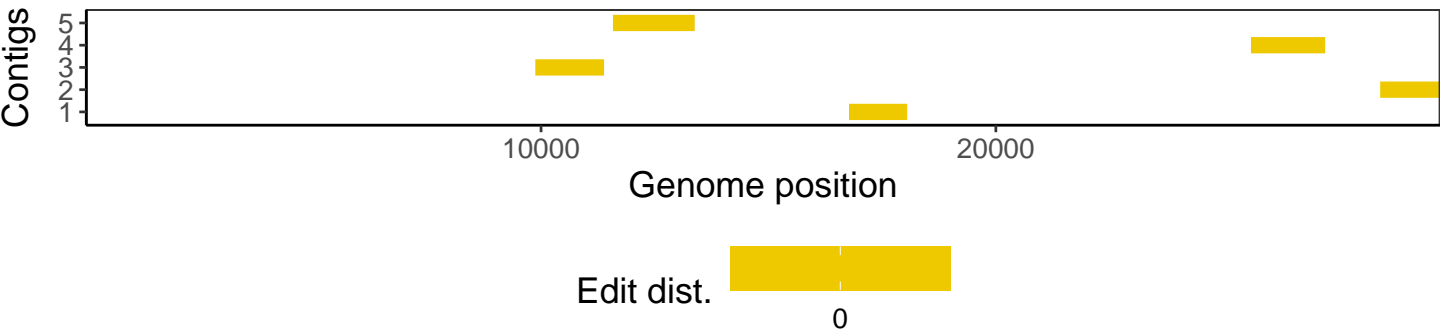
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



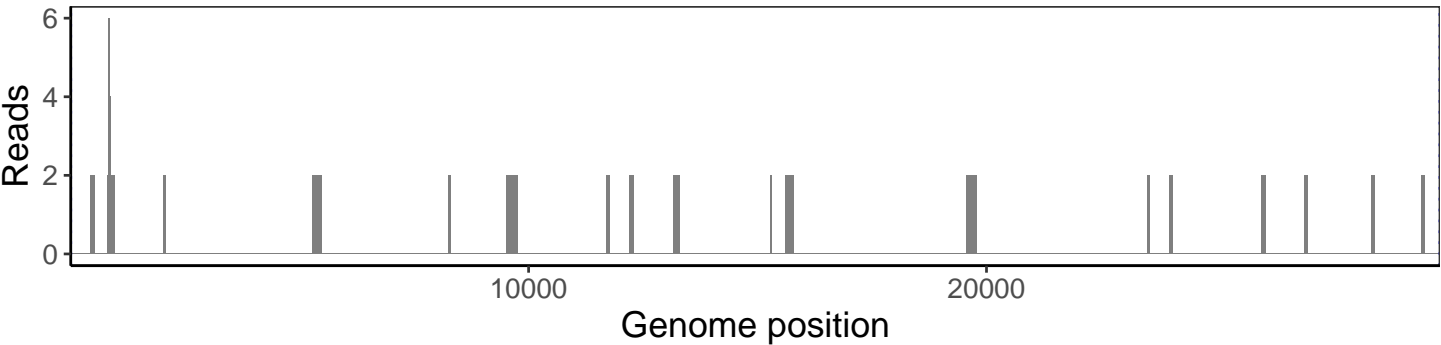
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



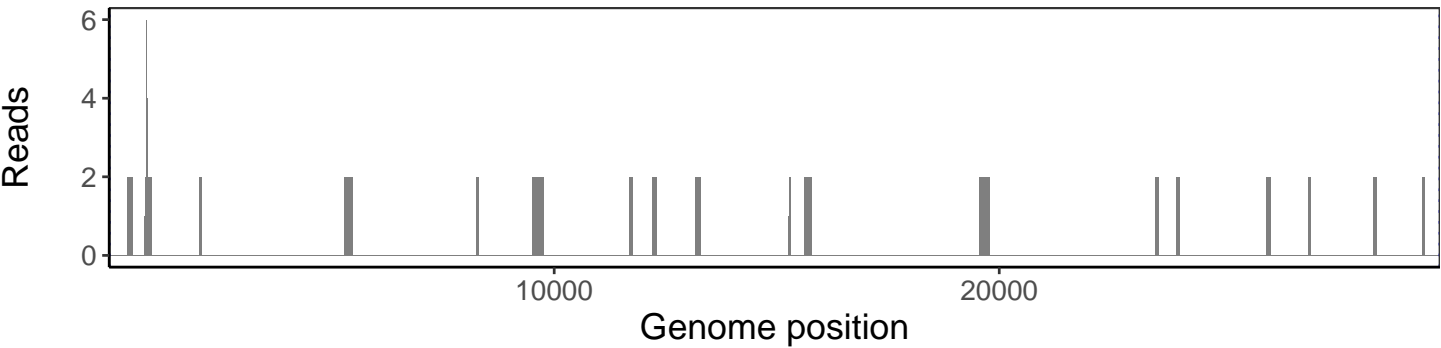
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



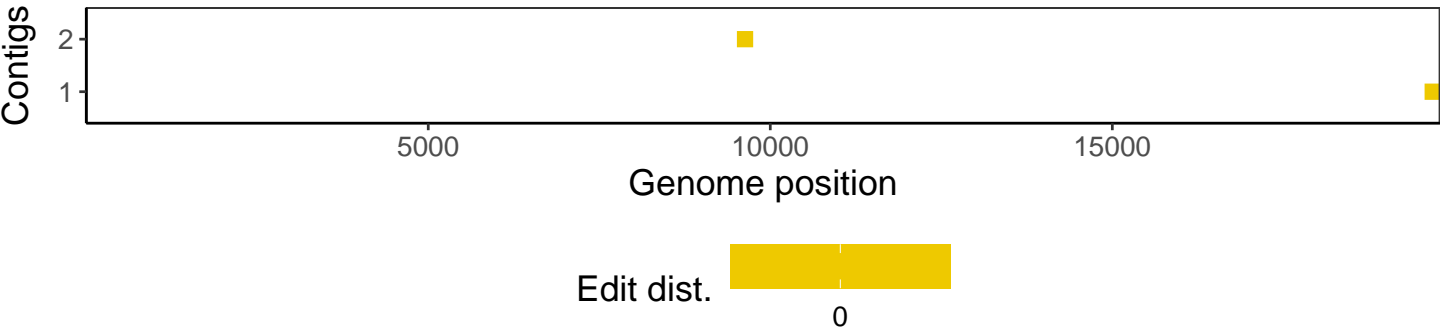
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



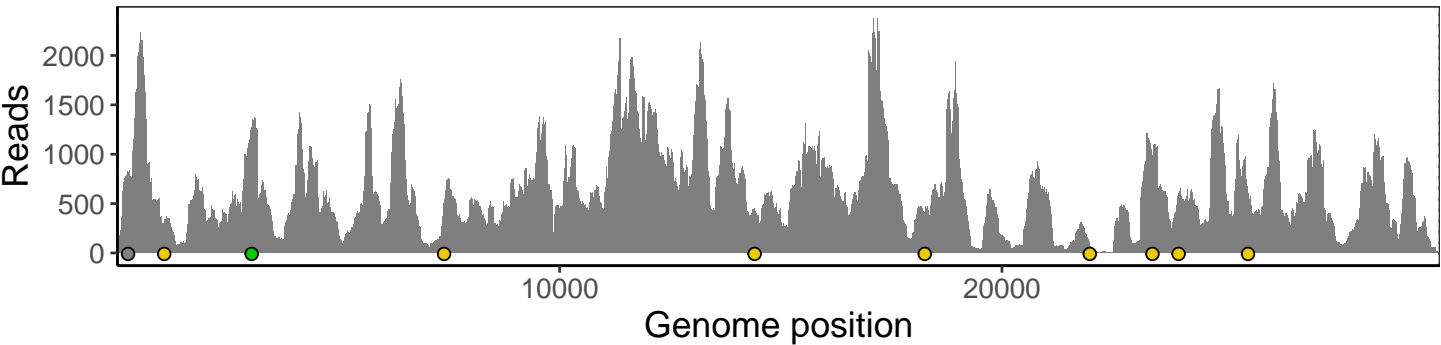
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



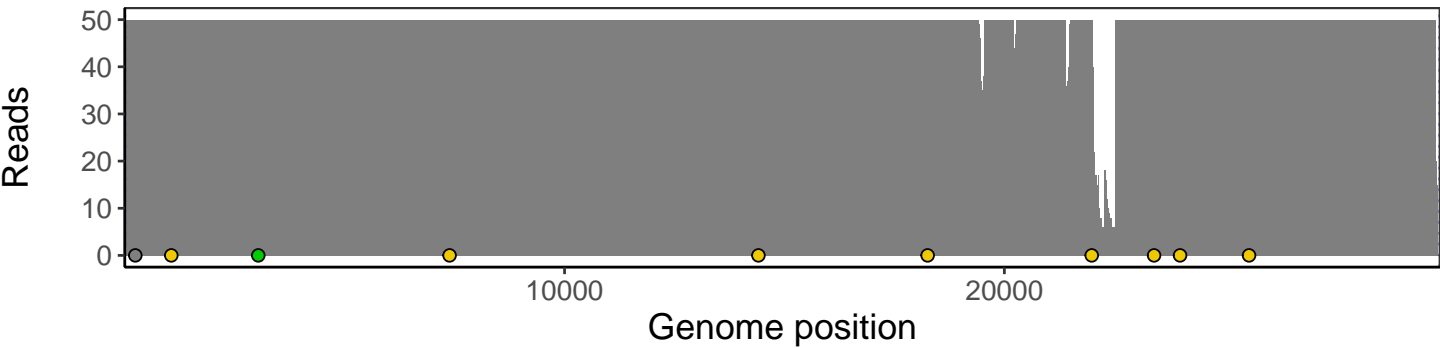
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



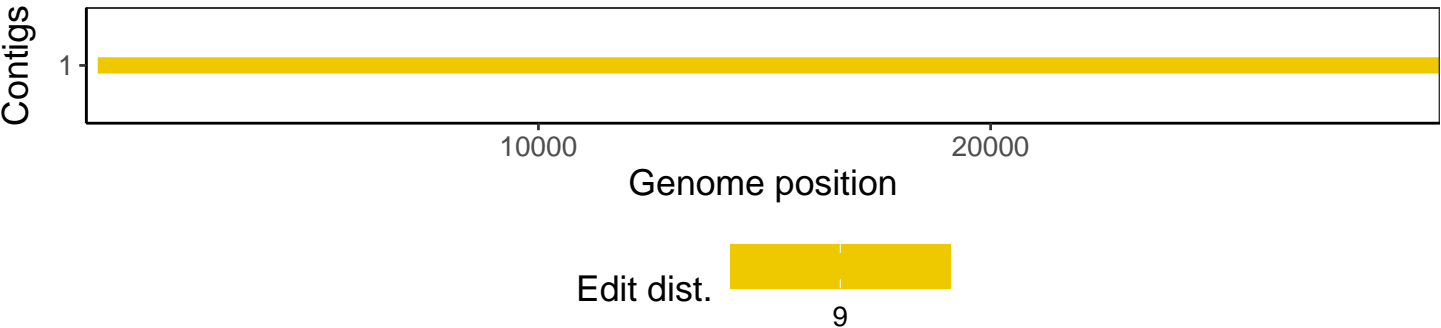
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



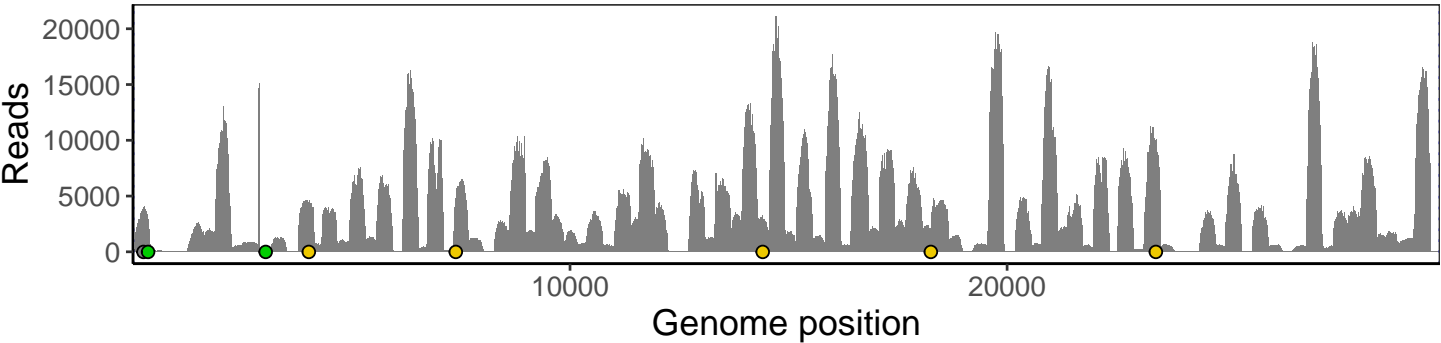
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



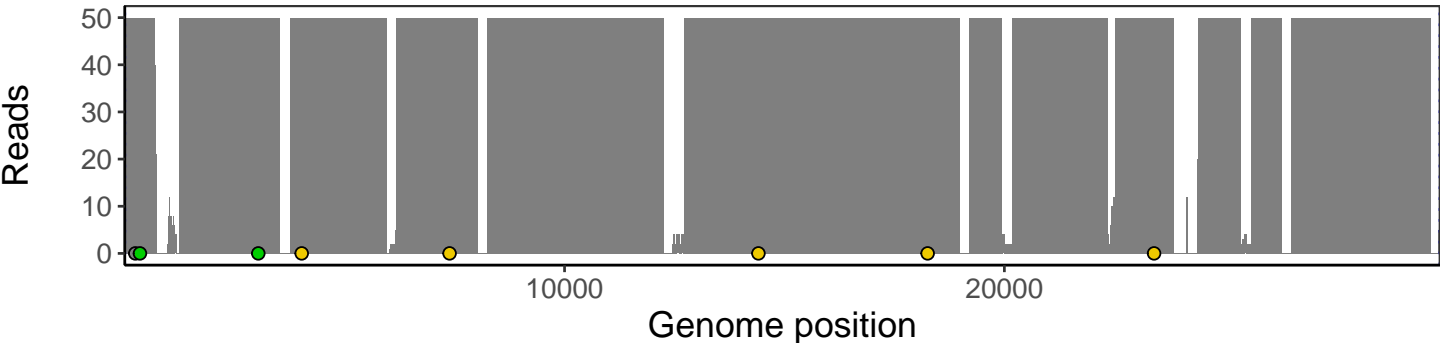
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



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.



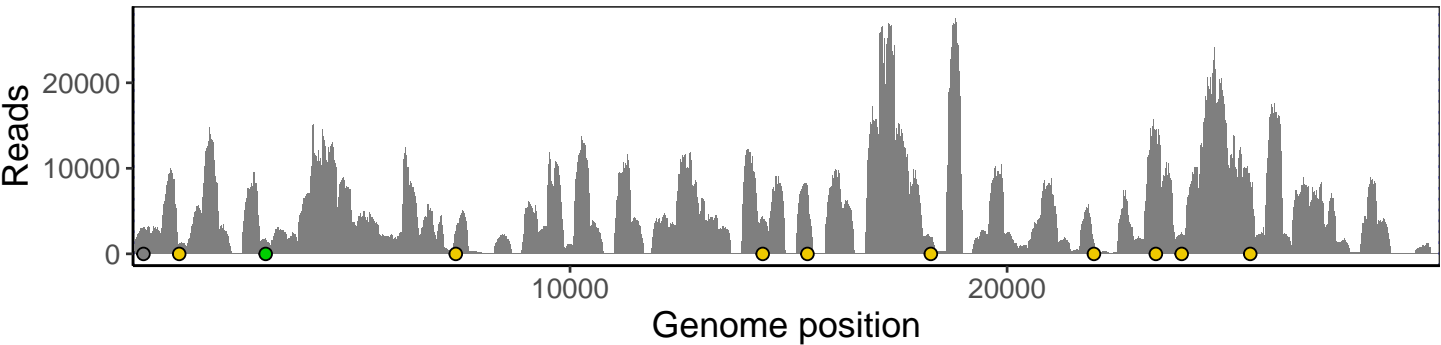
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



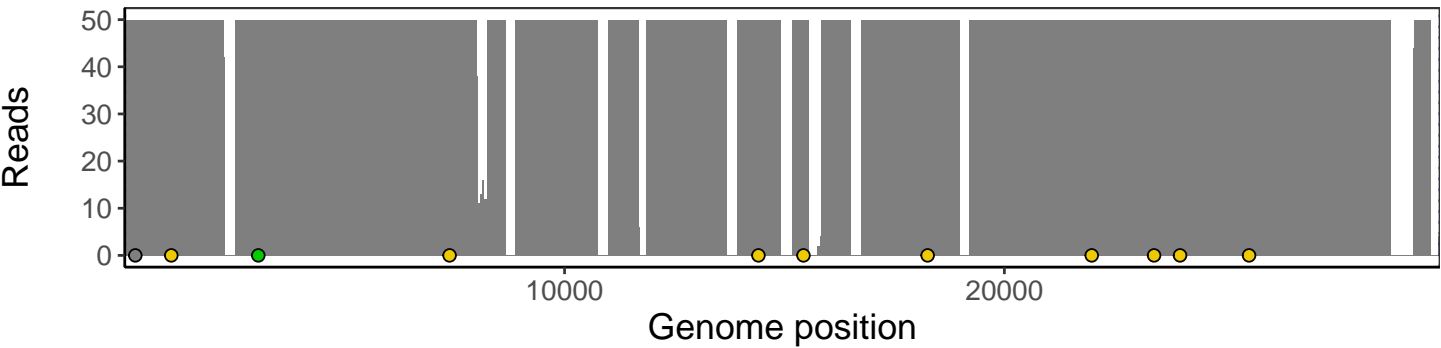
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



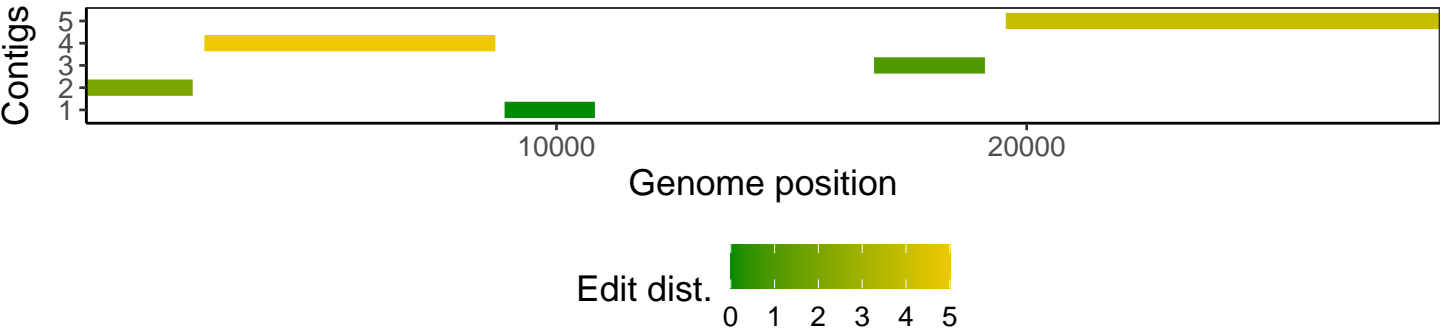
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



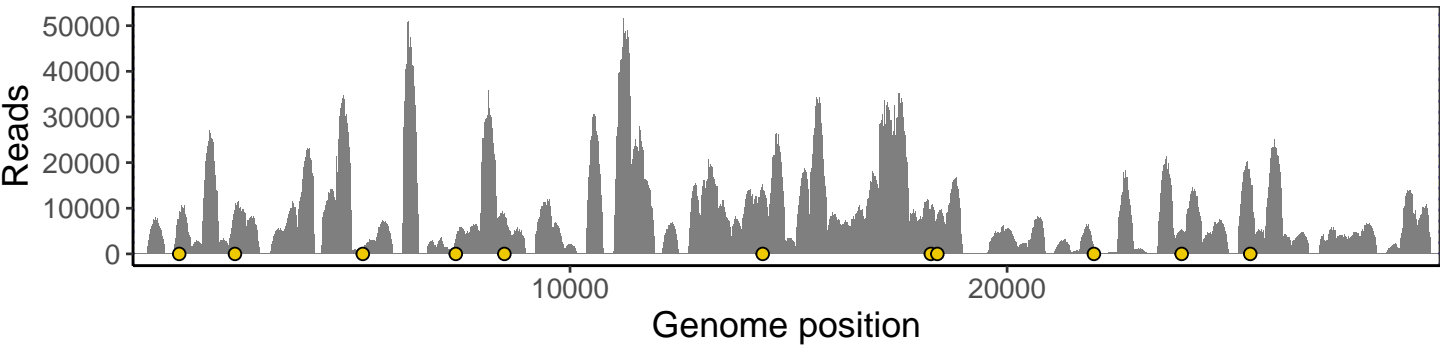
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



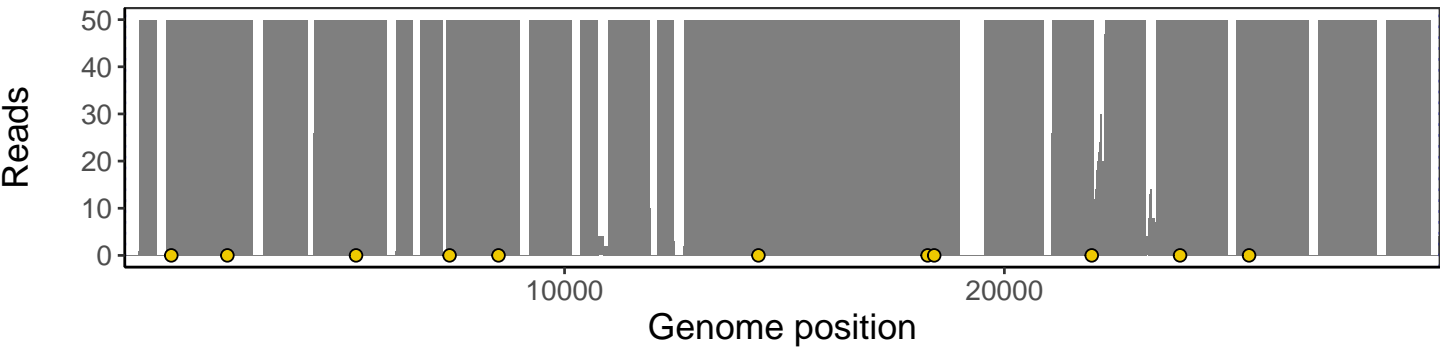
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



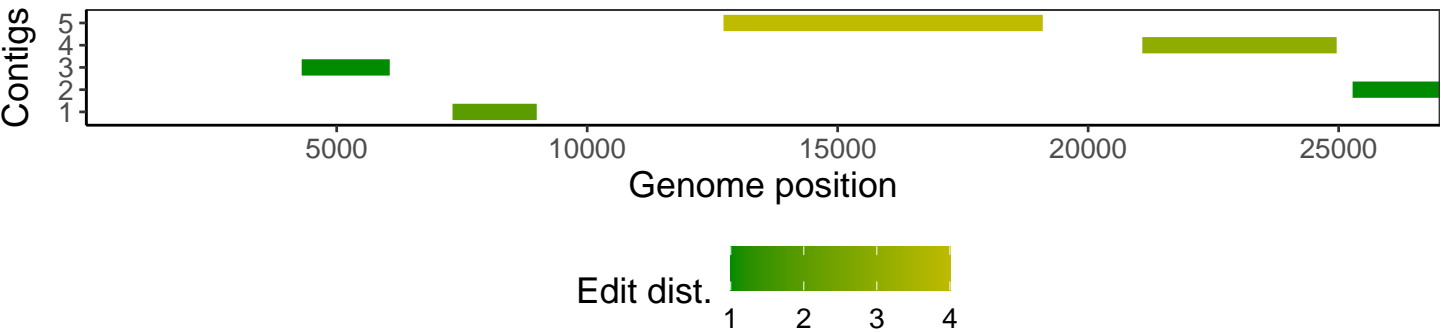
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

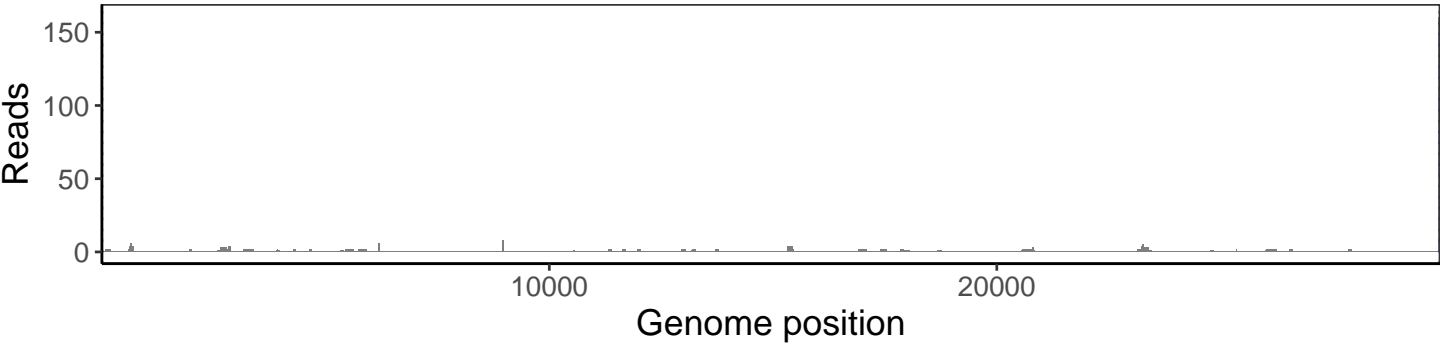


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

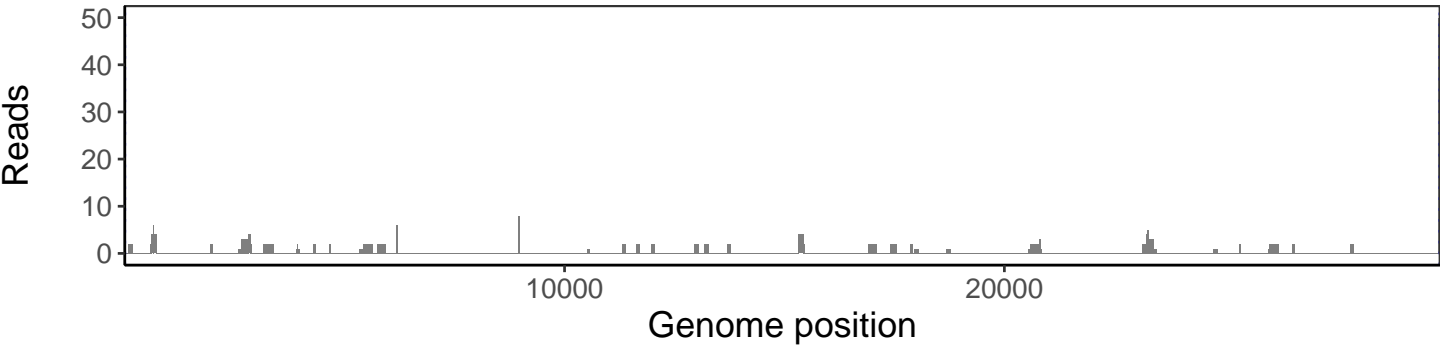




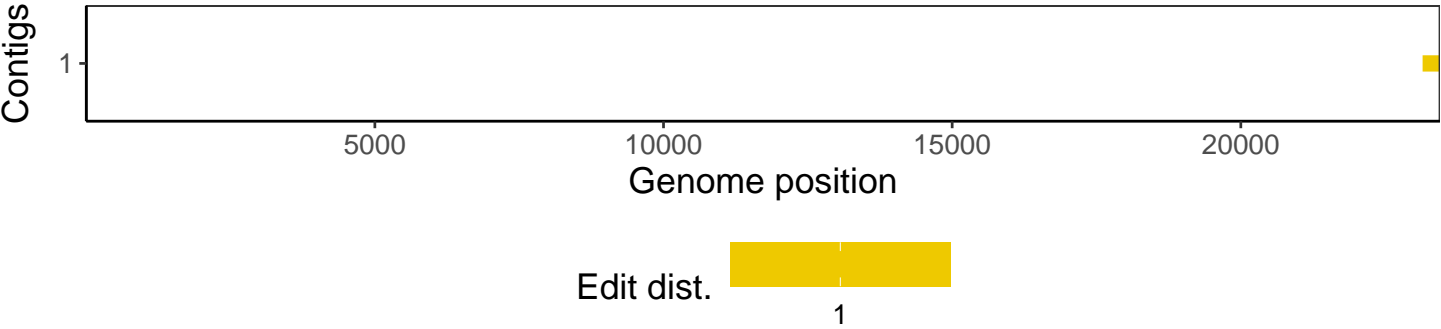
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



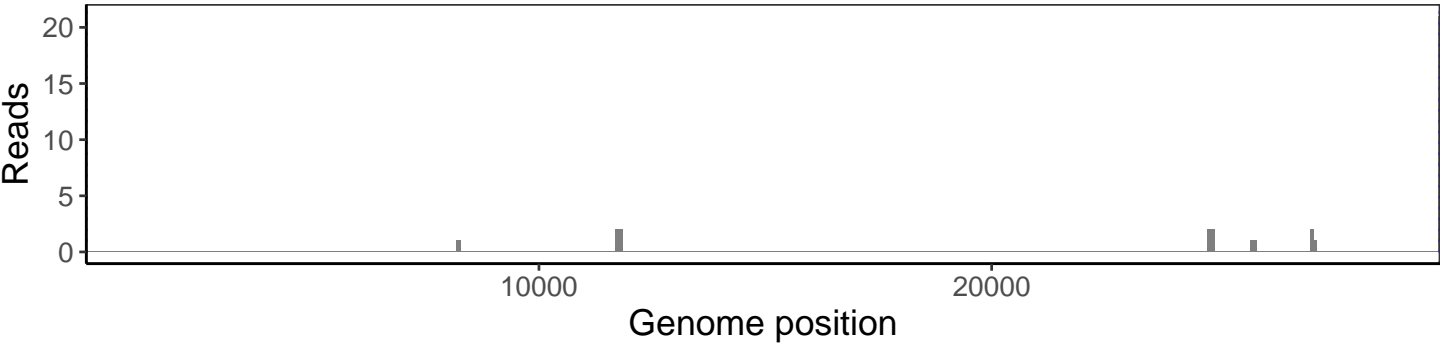
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



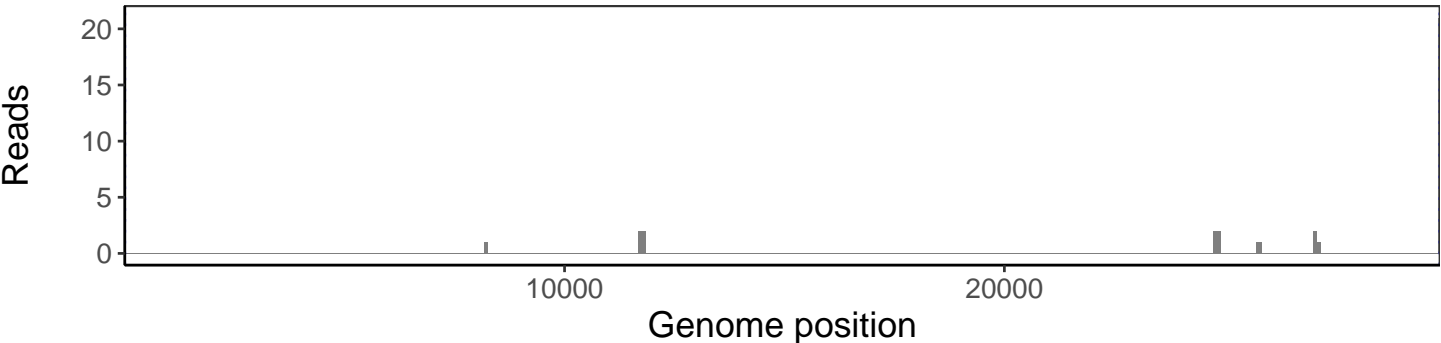
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



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.

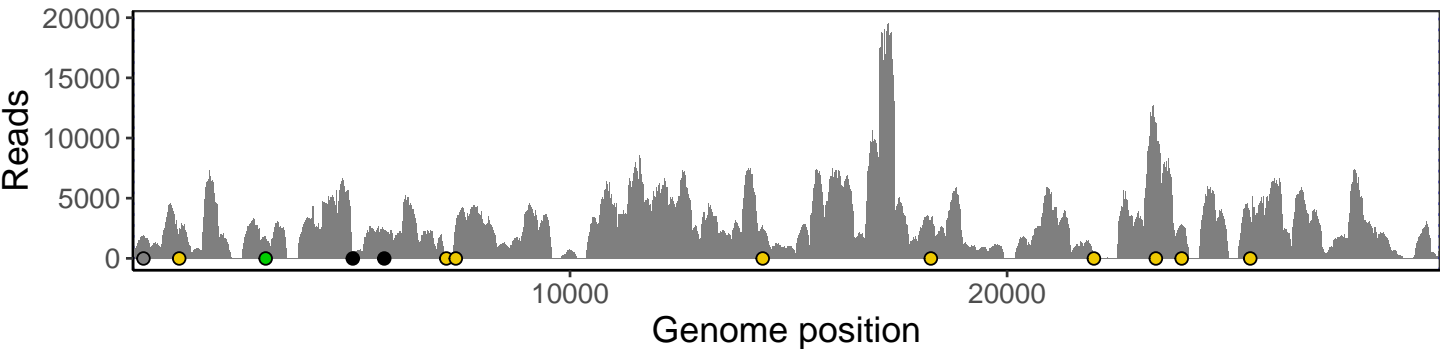


Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

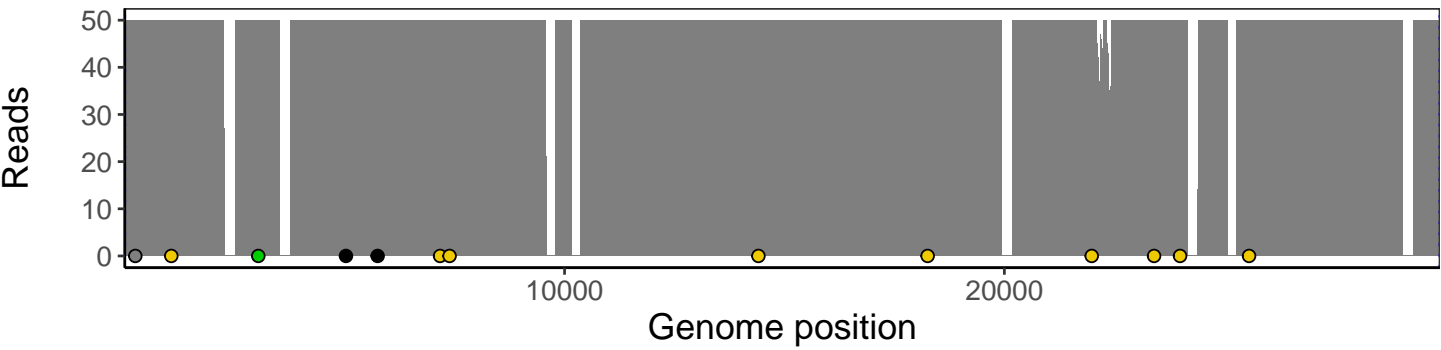


No contig data available.

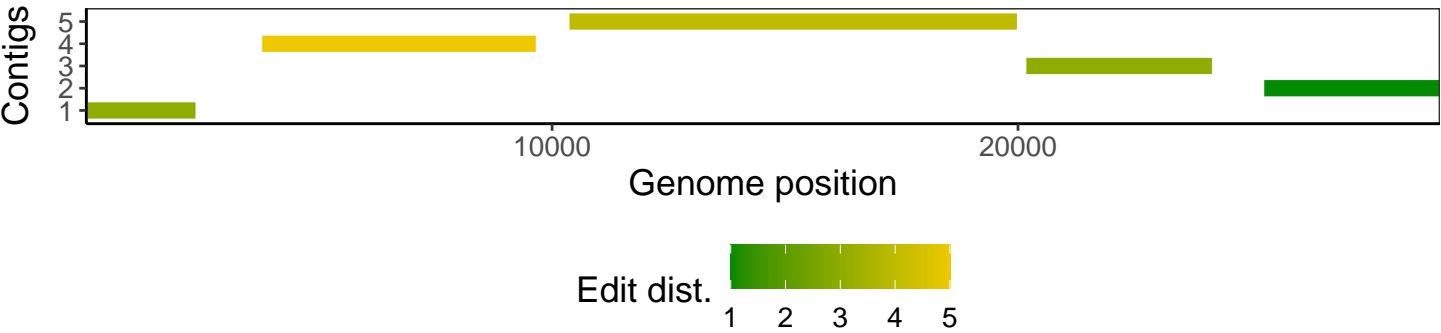
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



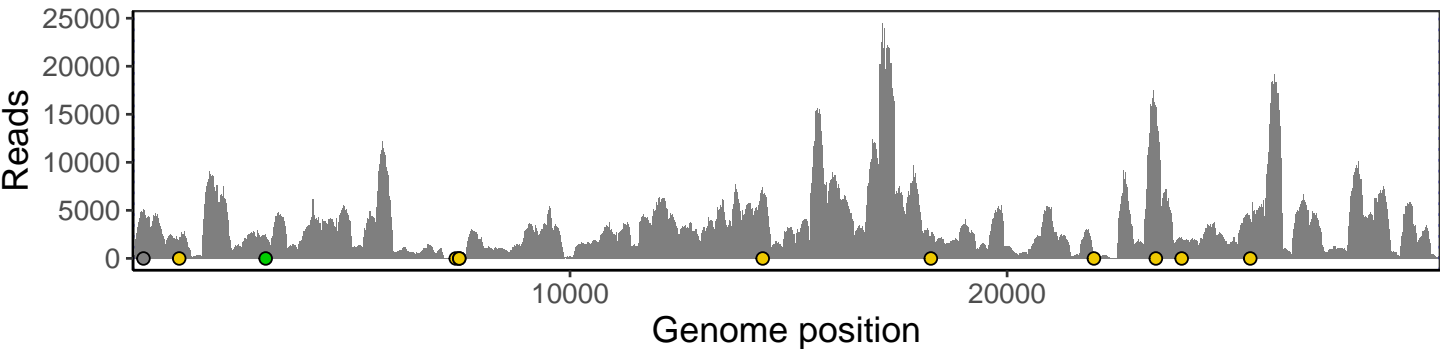
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



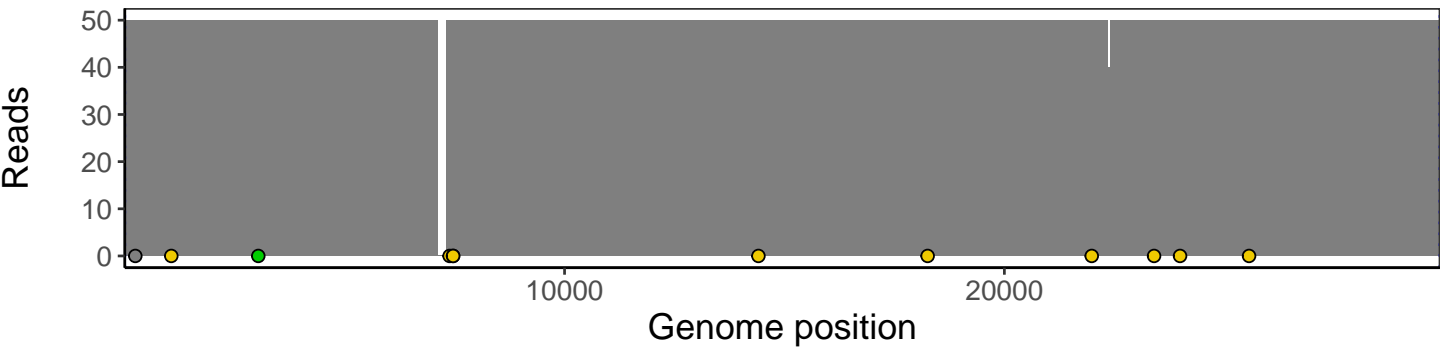
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



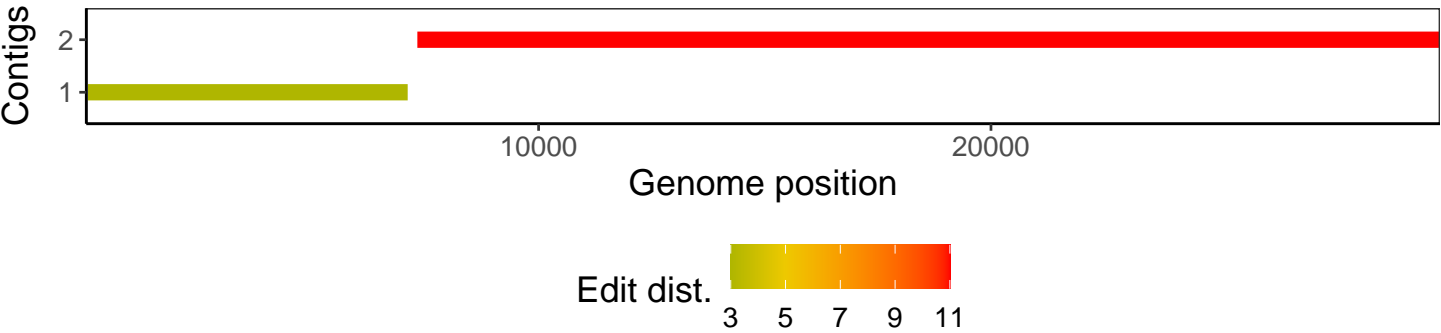
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



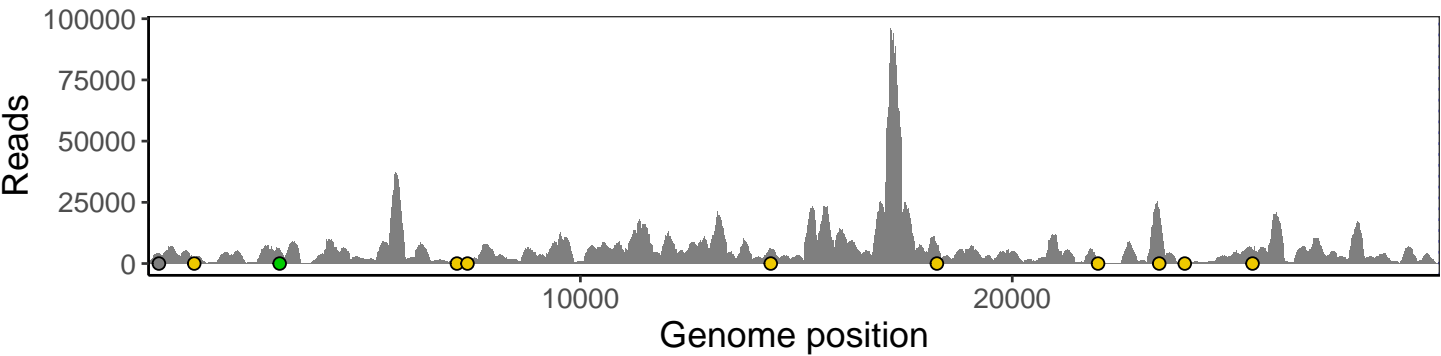
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



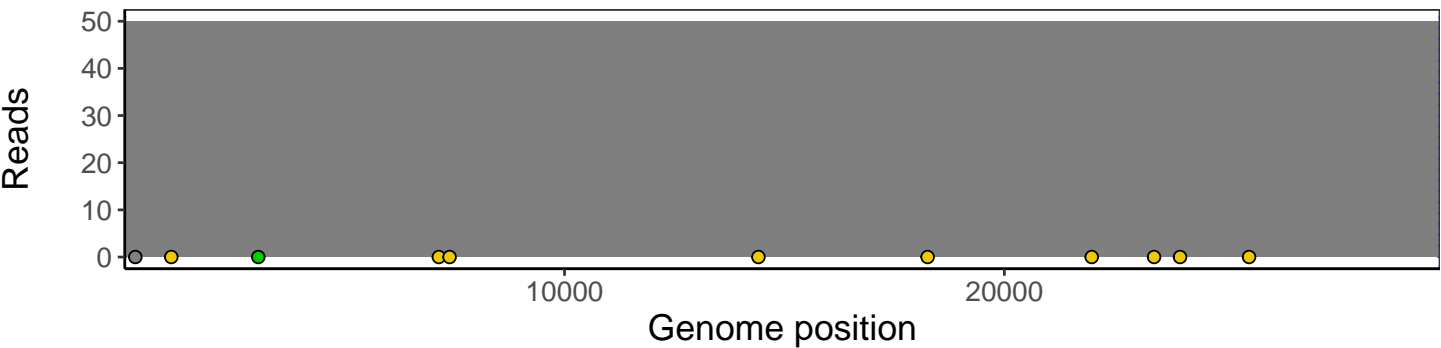
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



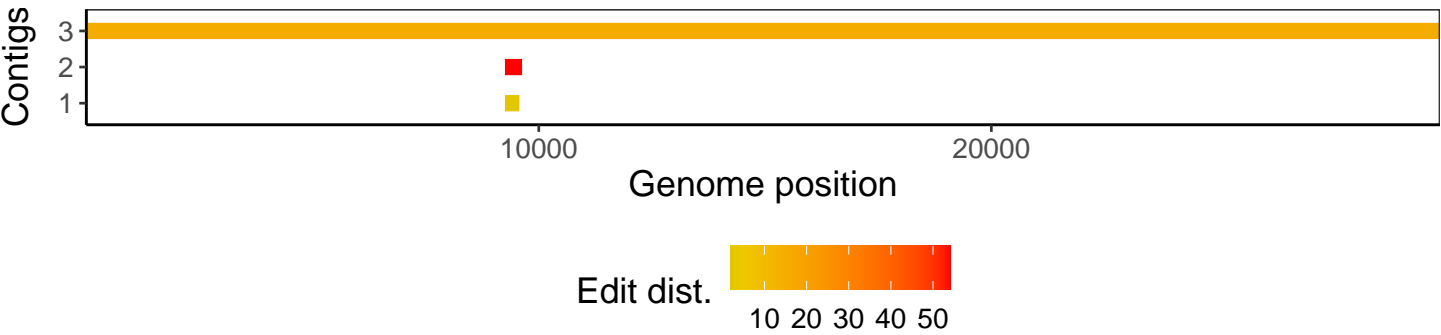
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



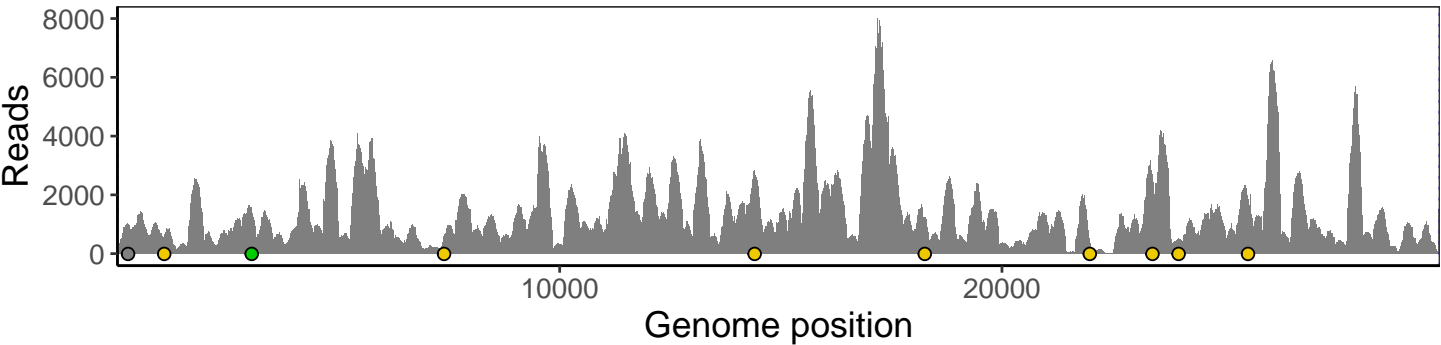
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



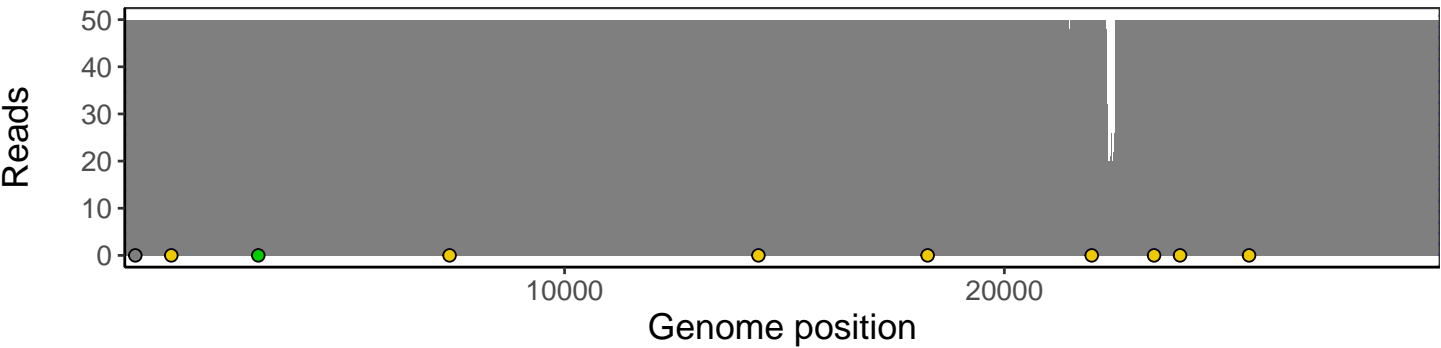
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



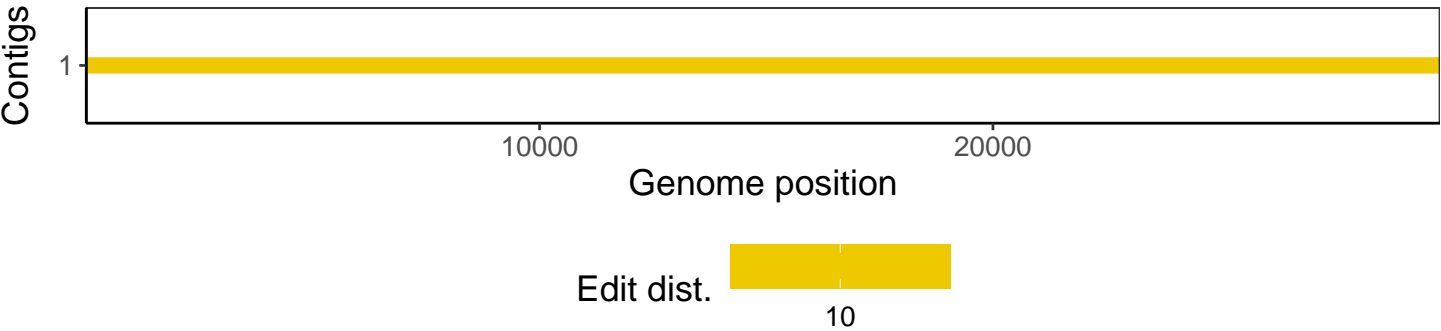
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	3.1.3
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	3.3.3
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