

COVID-19 subject 256

2021-04-30

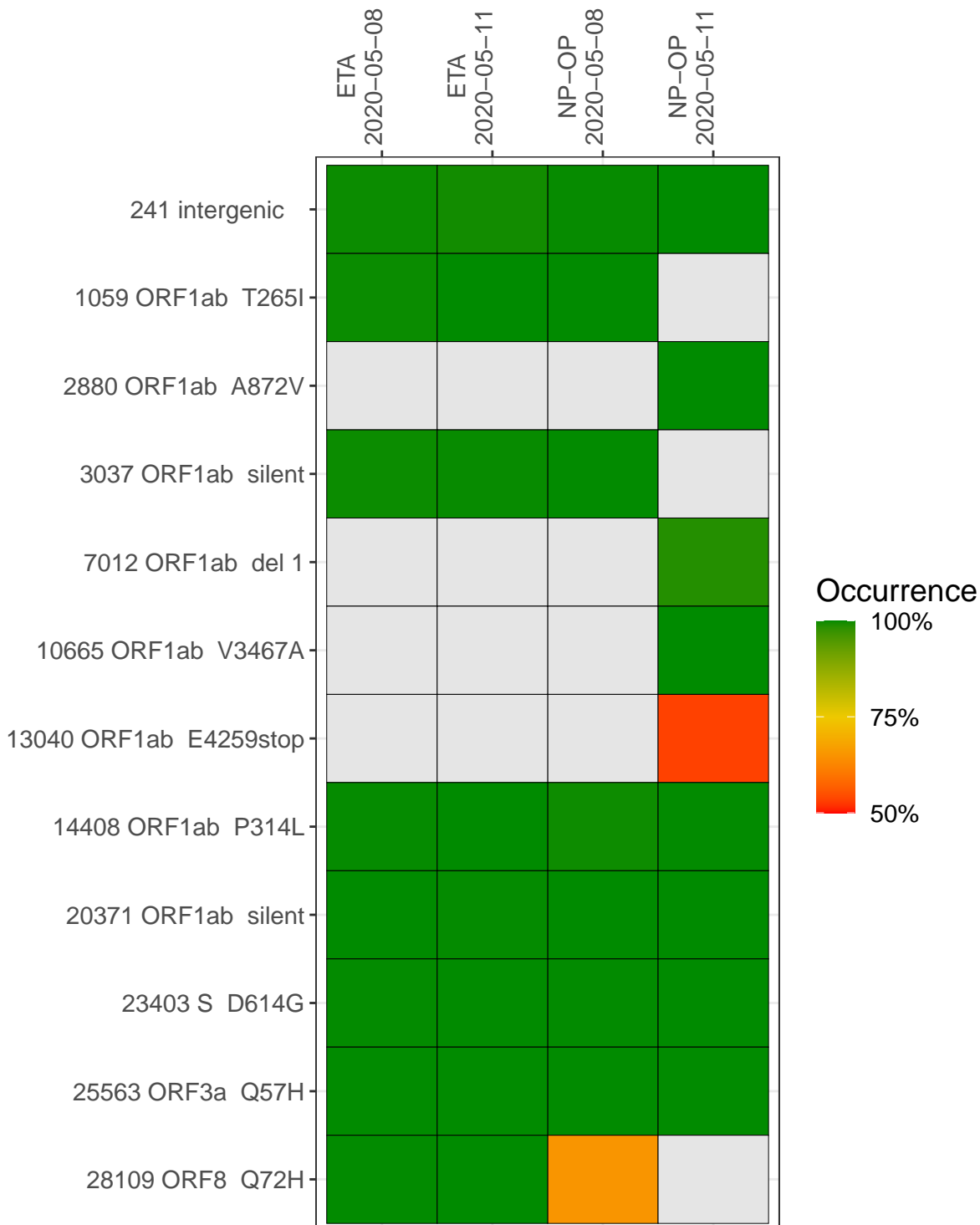
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP0107	composite	NA	NP-OP	2020-05-08	29.68	B.1	100.0%	99.9%
VSP0118	composite	NA	NP-OP	2020-05-11	2.63	NA	62.6%	58.9%
VSP0100-1	single experiment	2760000	ETA	2020-05-08	29.82	B.1	99.8%	99.8%
VSP0107-1	single experiment	595000	NP-OP	2020-05-08	29.89	B.1	99.9%	99.9%
VSP0107-2	single experiment	595000	NP-OP	2020-05-08	29.68	B.1	99.9%	99.8%
VSP0118-1	single experiment	269	NP-OP	2020-05-11	2.67	NA	56.7%	54.3%
VSP0118-2	single experiment	1345	NP-OP	2020-05-11	0.60	NA	9.4%	6.2%
VSP0118-3	single experiment	1345	NP-OP	2020-05-11	0.56	NA	11.2%	8.2%
VSP0118-4	single experiment	1345	NP-OP	2020-05-11	0.60	NA	11.1%	7.4%
VSP0123-1	single experiment	123000	ETA	2020-05-11	29.26	B.1	99.9%	99.8%

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.



		ETA 2020-05-08		ETA 2020-05-11		NP-OP 2020-05-08		NP-OP 2020-05-11			
3 ORF8 (ORF3a 03 S D6 ORF1ab ORF1ab ORF1ab ORF1ab ORF1ab ORF1ab ORF1ab ORF1ab intergen	intergen	1920	4157	10740	3966	1847	0	0	0		
	ORF1ab	4367	3378	12059	3388	0	0	0	0		
	ORF1ab	1461	2759	3527	5145	47	0	0	0		
	ORF1ab	770	2277	3930	5260	0	0	0	0		
	ORF1ab	40	2156	56	883	1655	0	0	0		
	ORF1ab	1181	4273	1592	2295	0	43	19	33		
	ORF1ab	17608	7548	15594	6663	5519	0	0	0		
	ORF1ab	1562	4511	6517	2547	4597	0	0	0		
	ORF1ab	55	2015	31	476	2092	0	0	0		
	ORF1ab	35115	10707	17343	12357	2425	0	0	2		
	ORF3a	4636	5697	10428	5419	2	81797	0	0		
	ORF8	4918	8559	8006	13510	0	0	0	0		
		VSP0100-1	VSP0123-1	VSP0107-1	VSP0107-2	VSP0118-1	VSP0118-2	VSP0118-3	VSP0118-4		

Base change

Expected

A

T

C

G

N

Ins/Del

No data

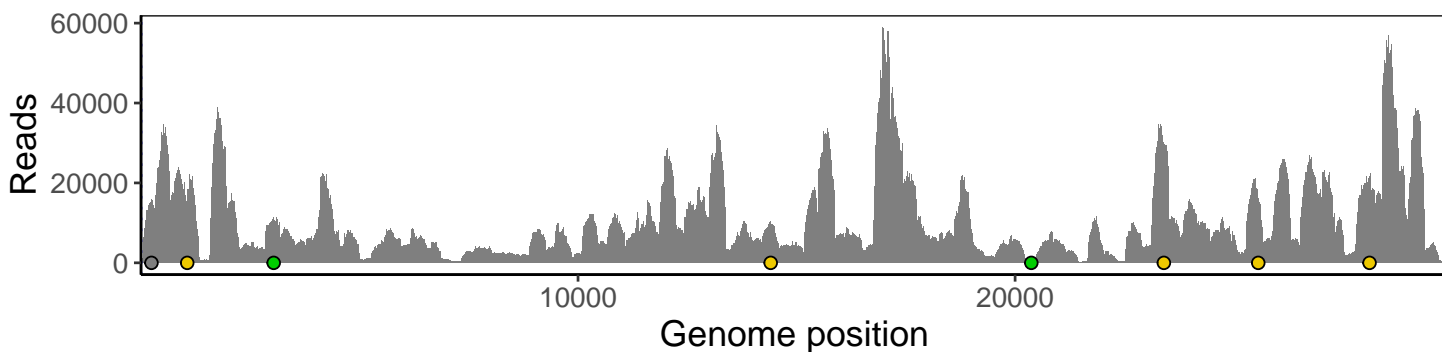
Base change

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

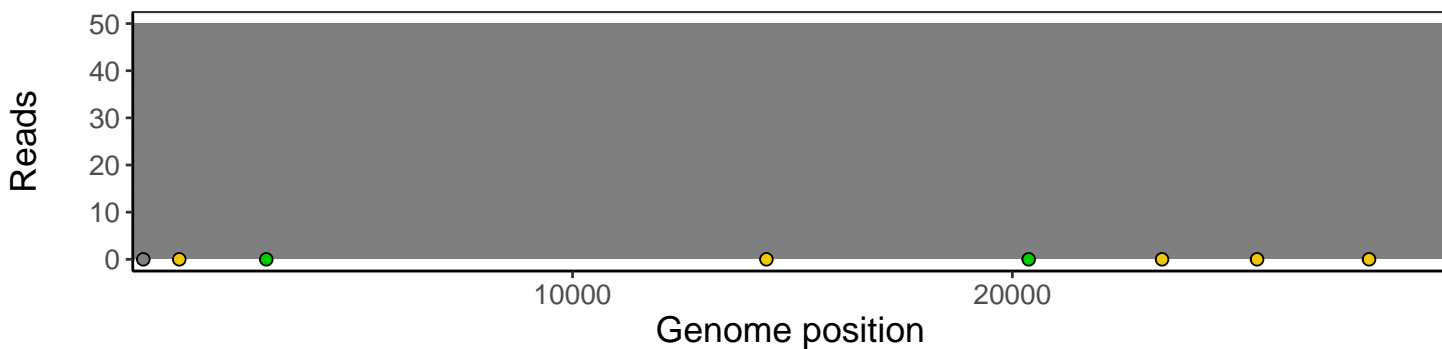
Analyses of individual experiments and composite results

VSP0107 | 2020-05-08 | NP-OP | 256no-t | composite result

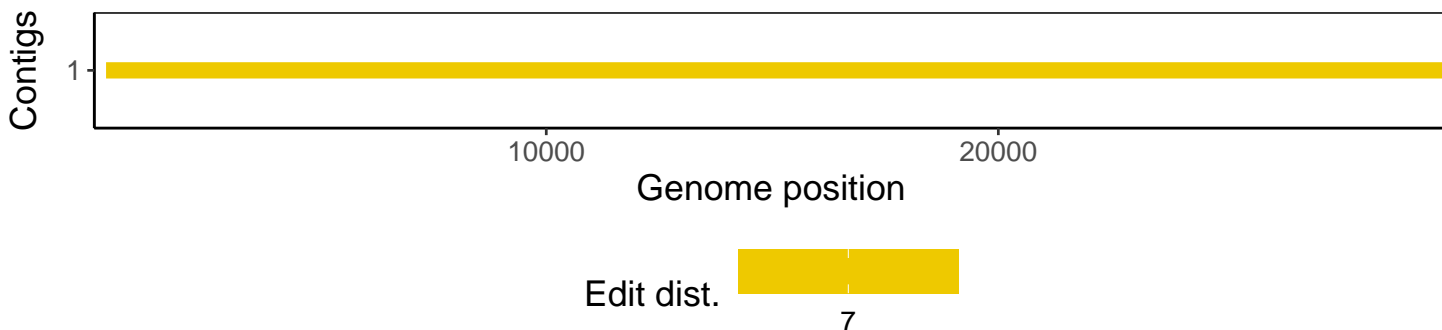
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



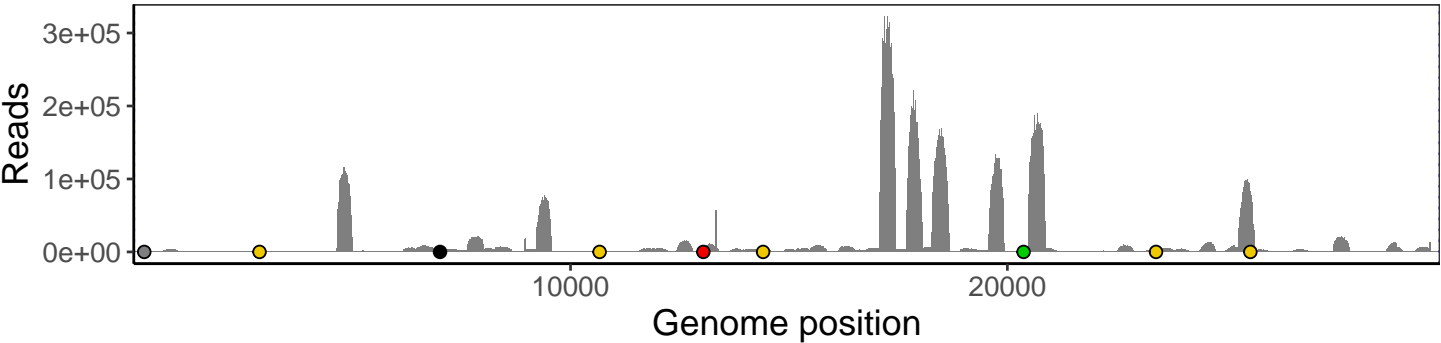
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



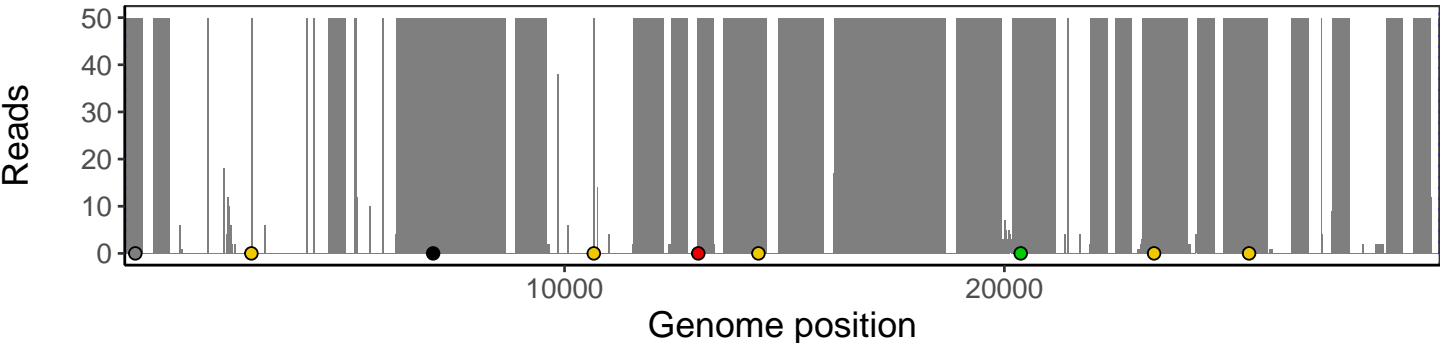
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



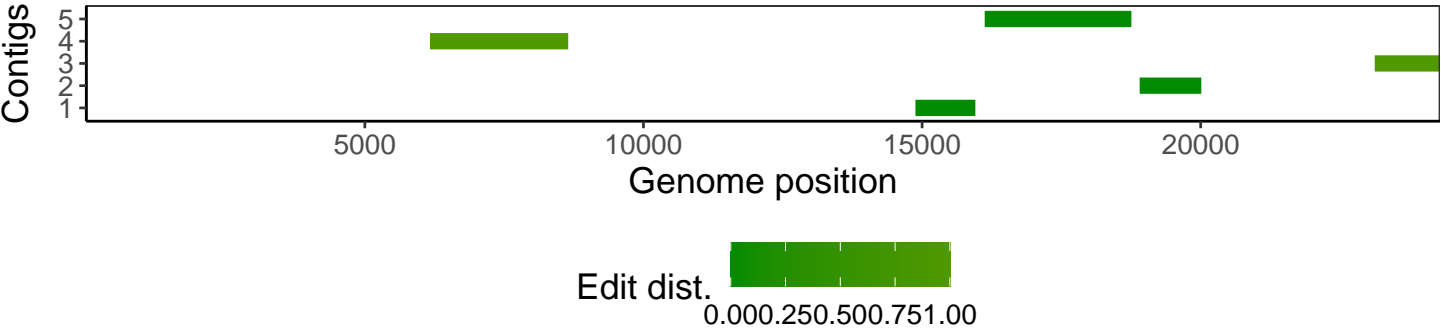
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



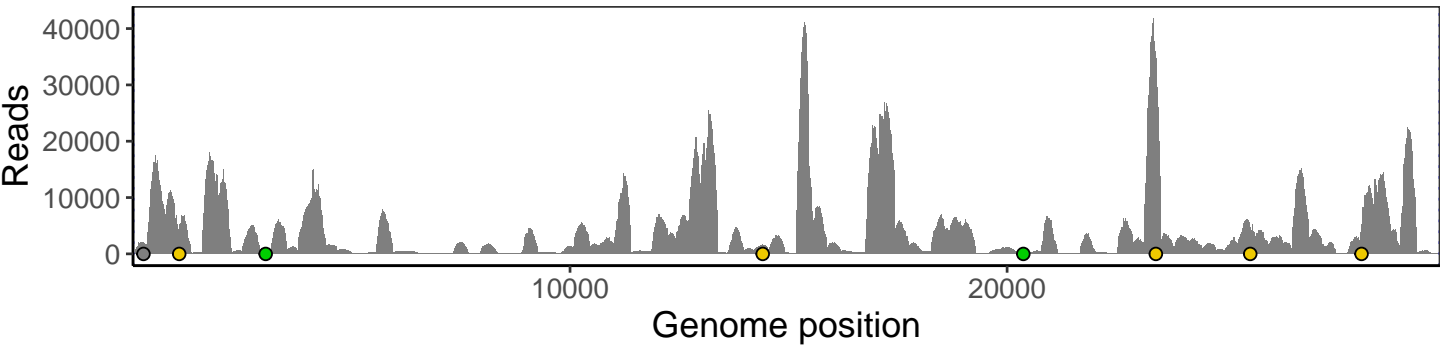
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



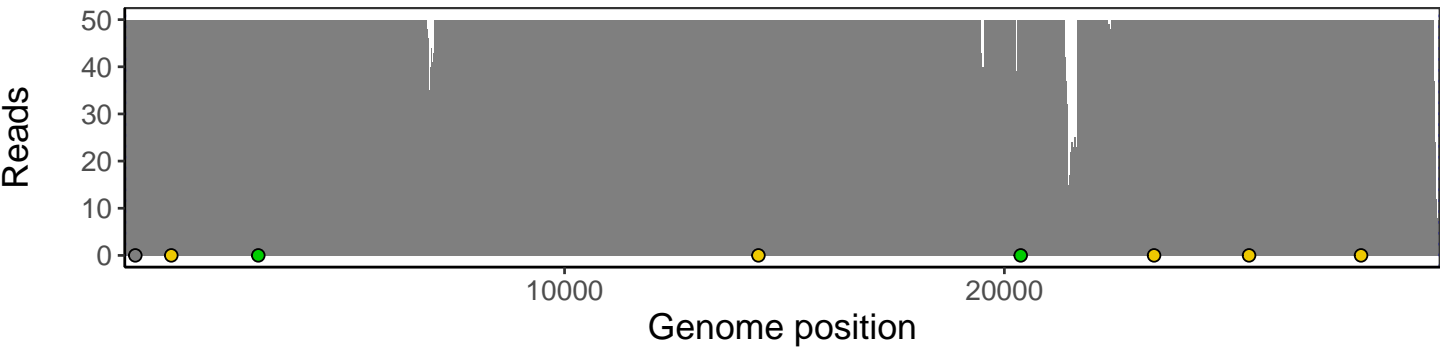
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



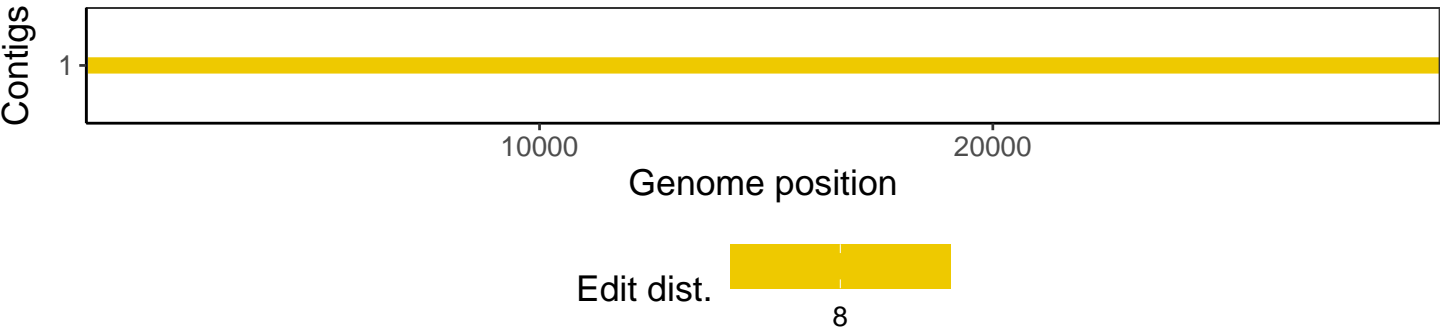
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



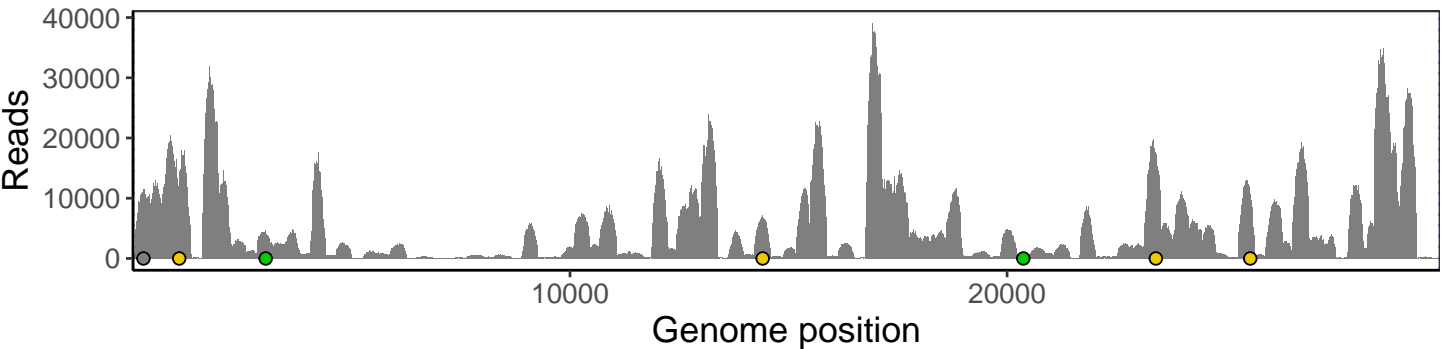
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



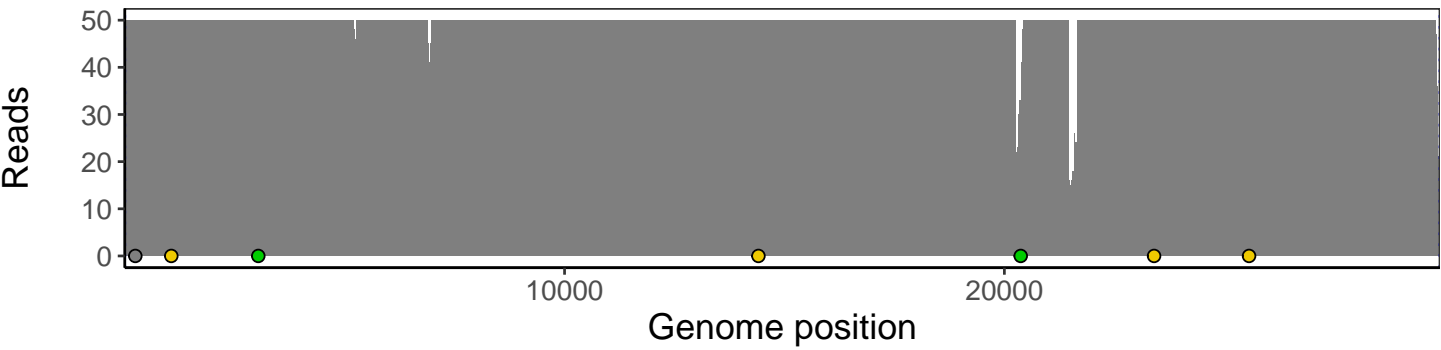
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



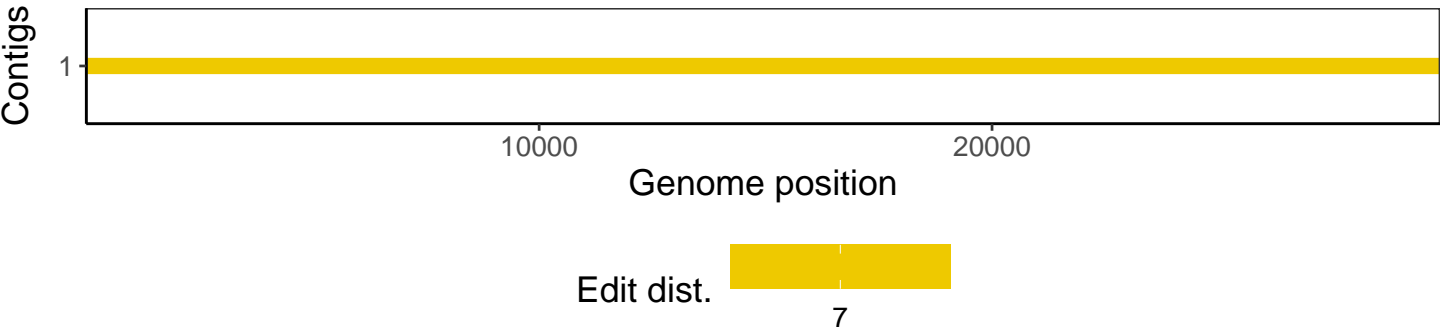
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



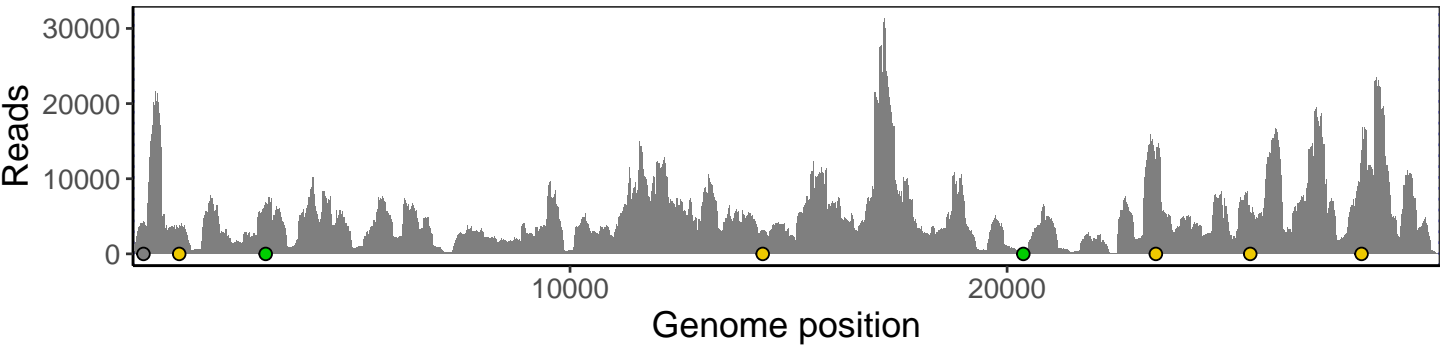
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



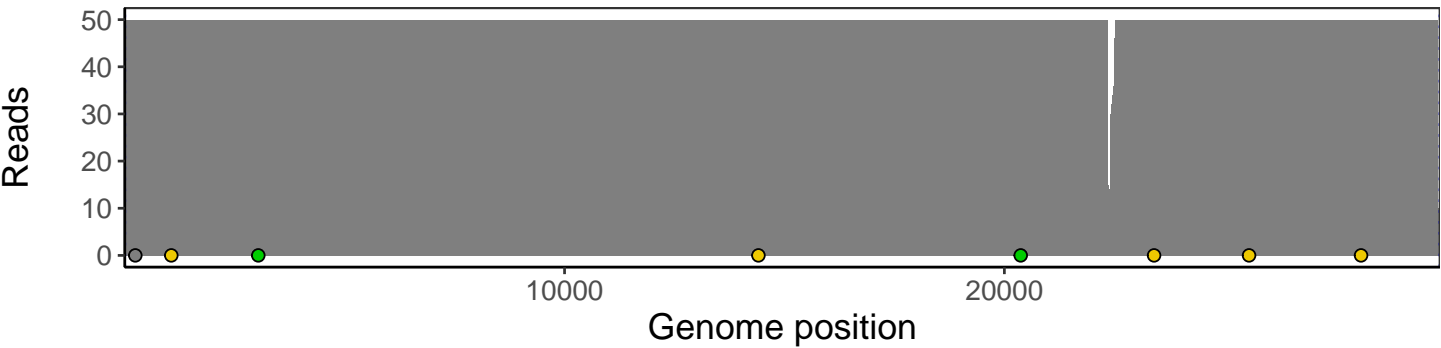
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



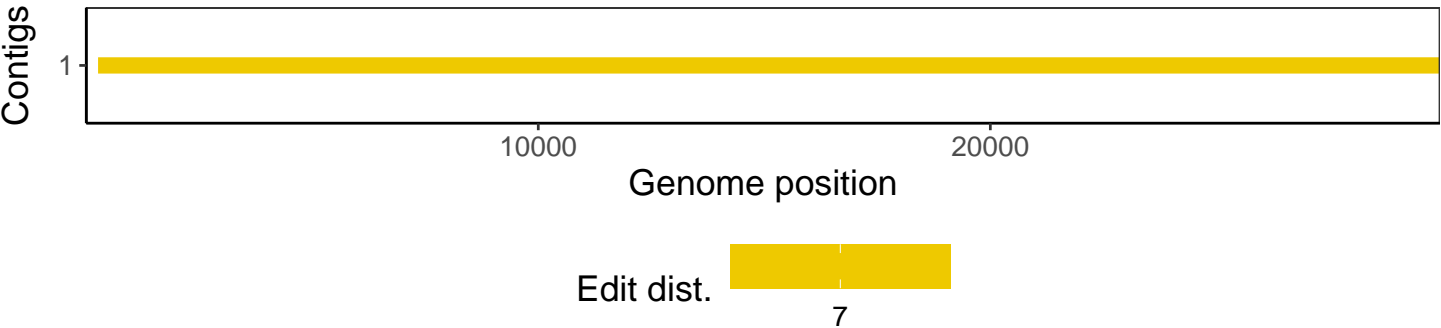
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



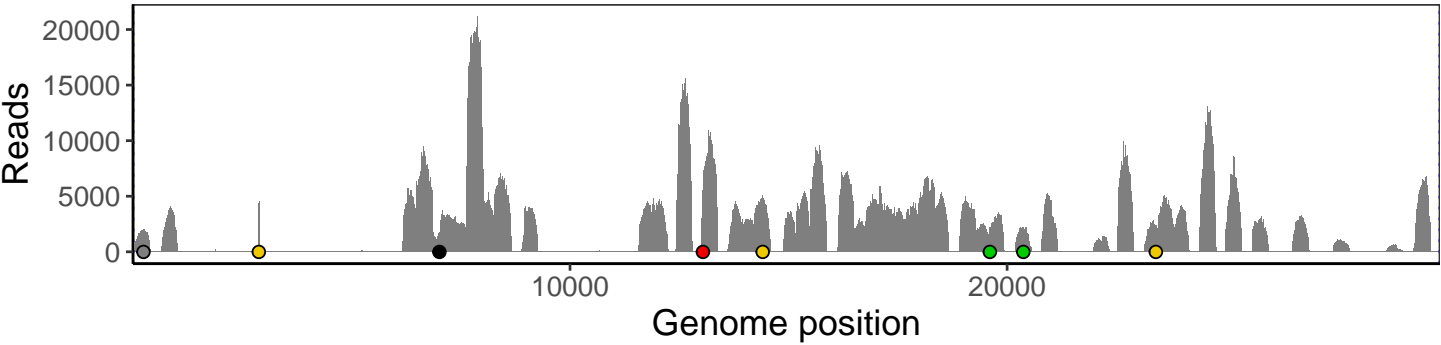
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



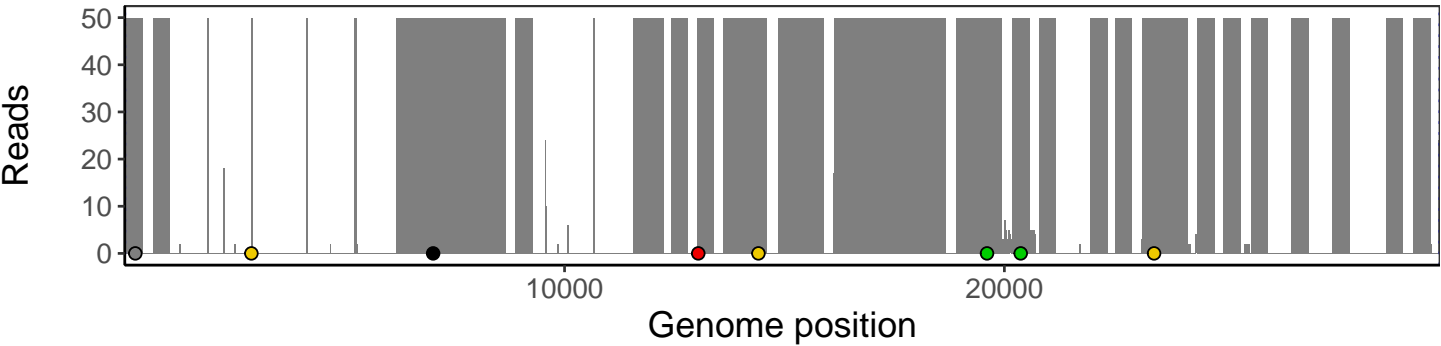
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



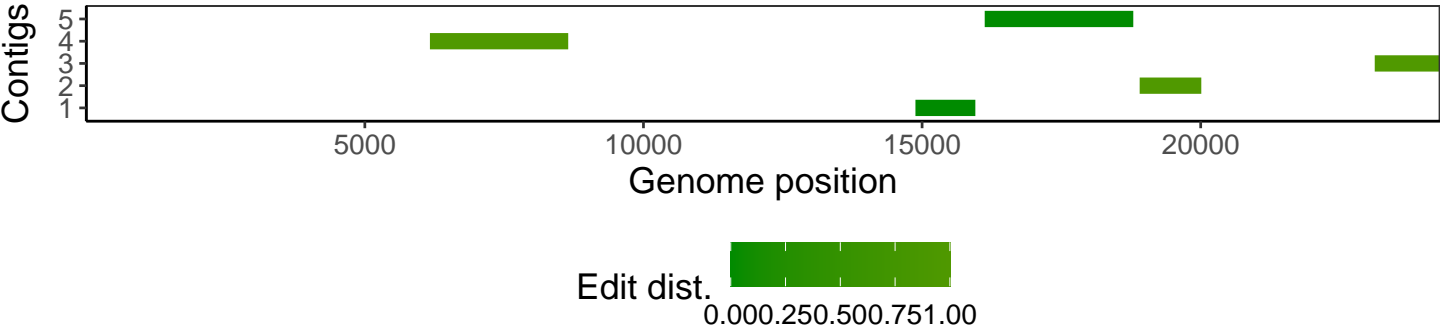
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



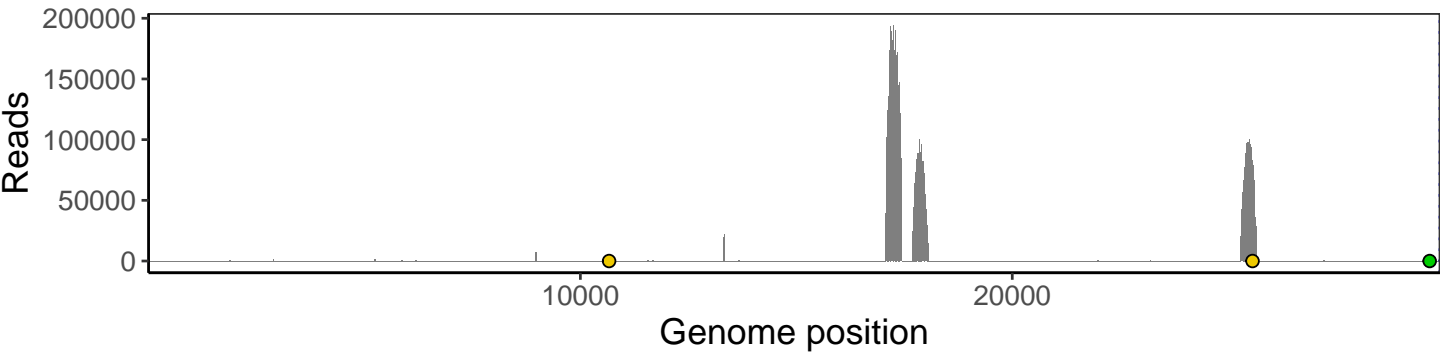
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



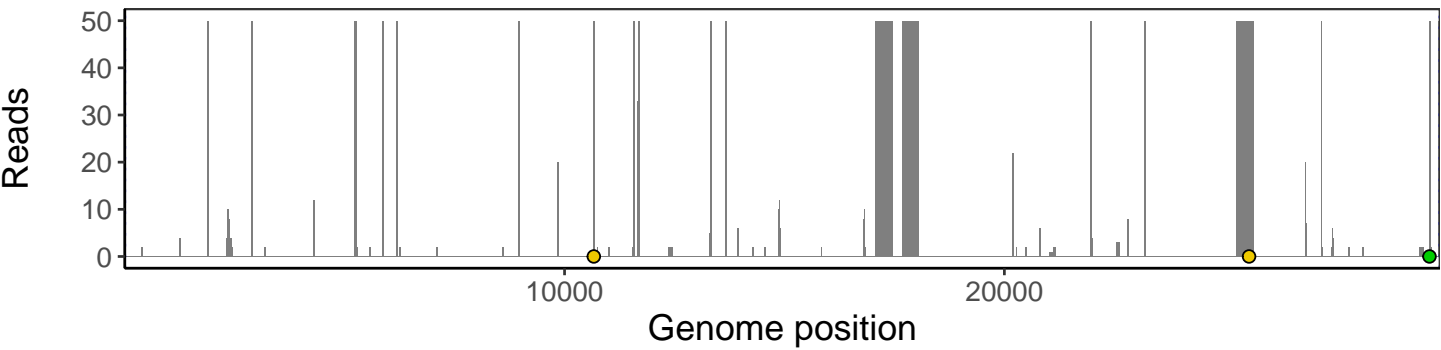
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



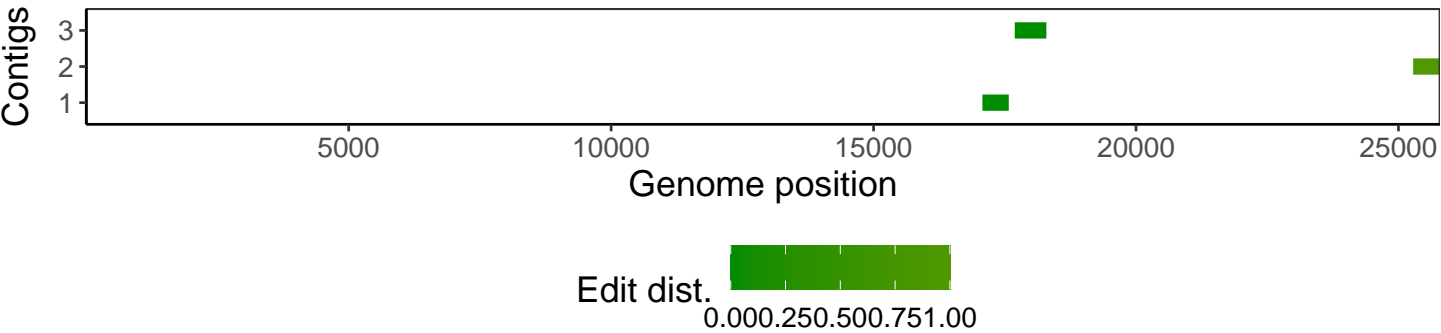
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



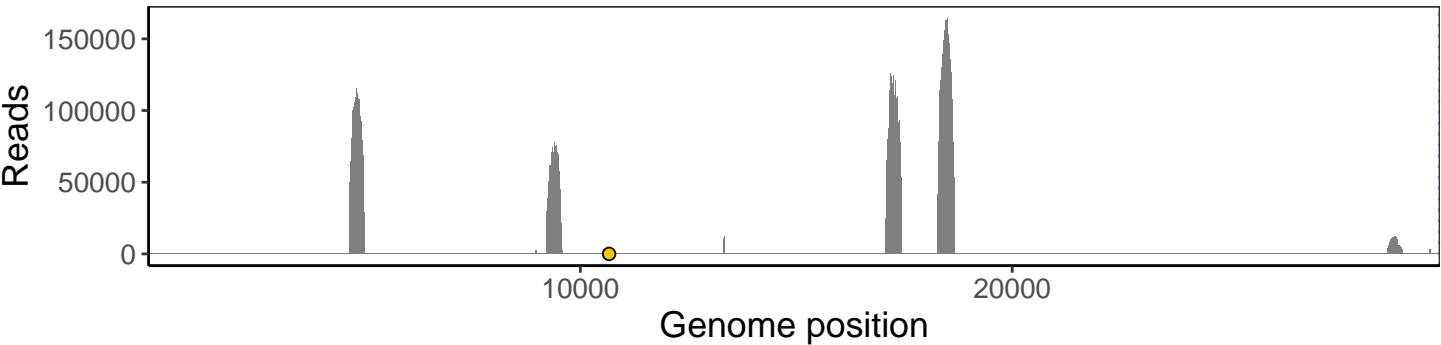
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



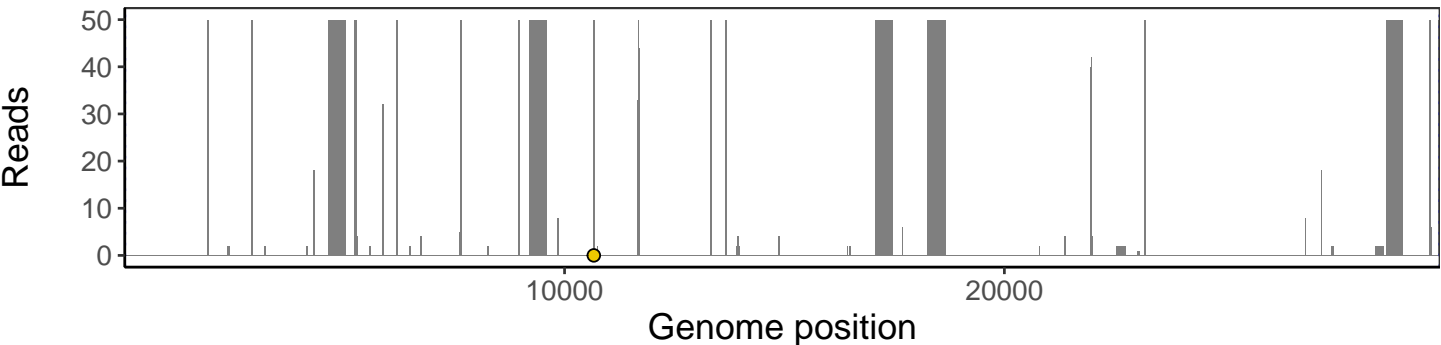
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



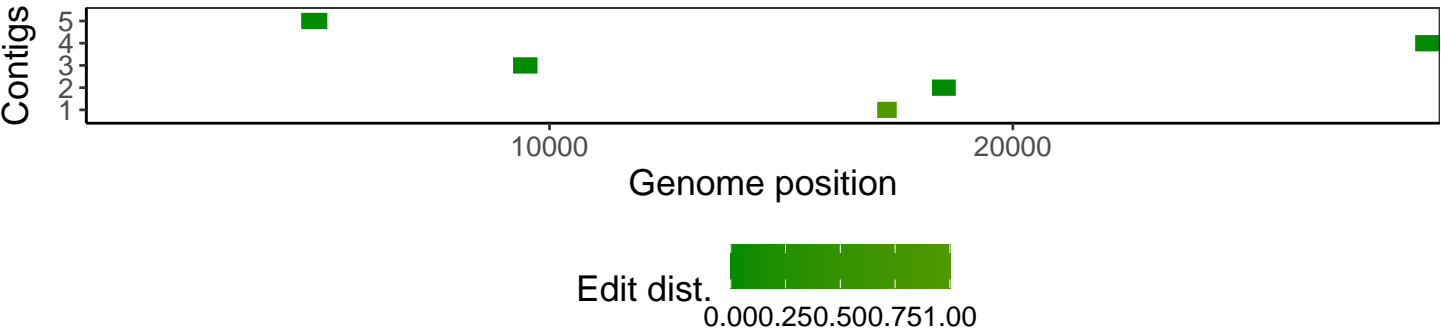
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



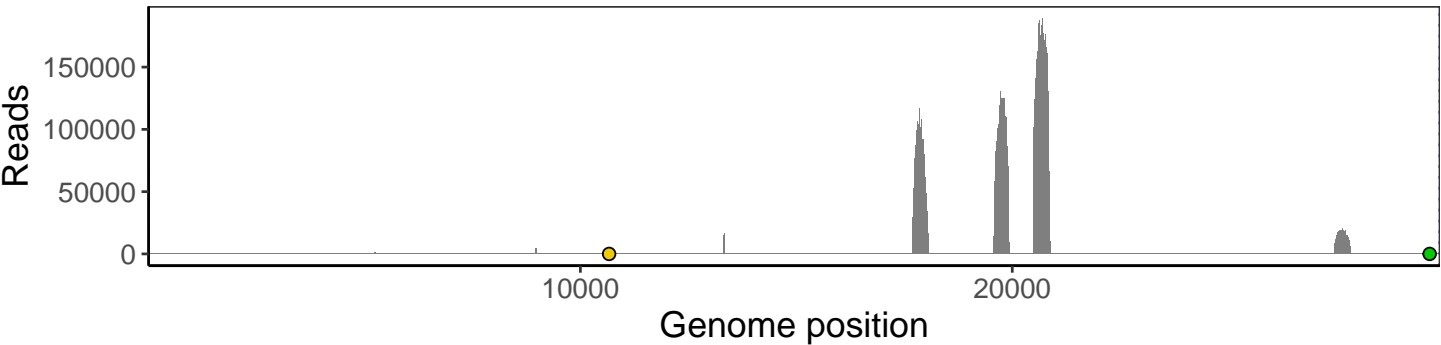
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



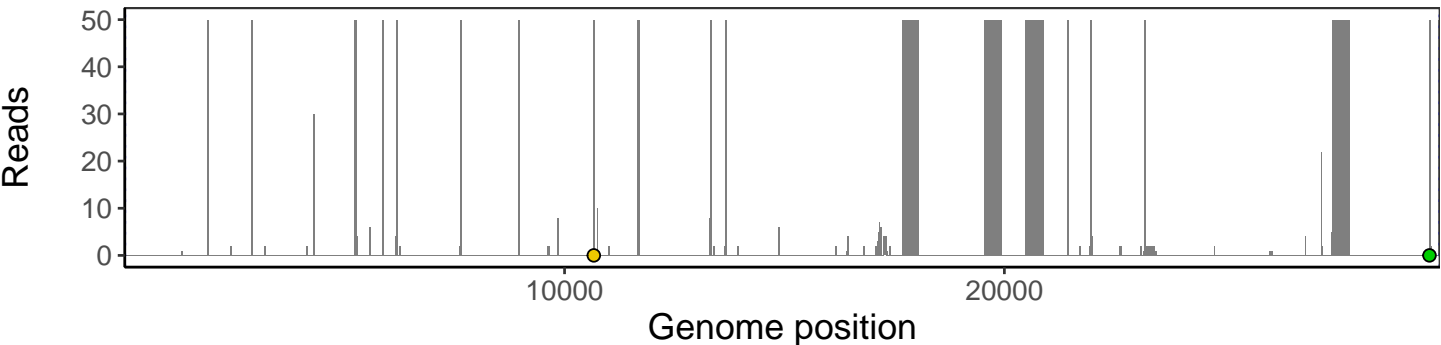
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



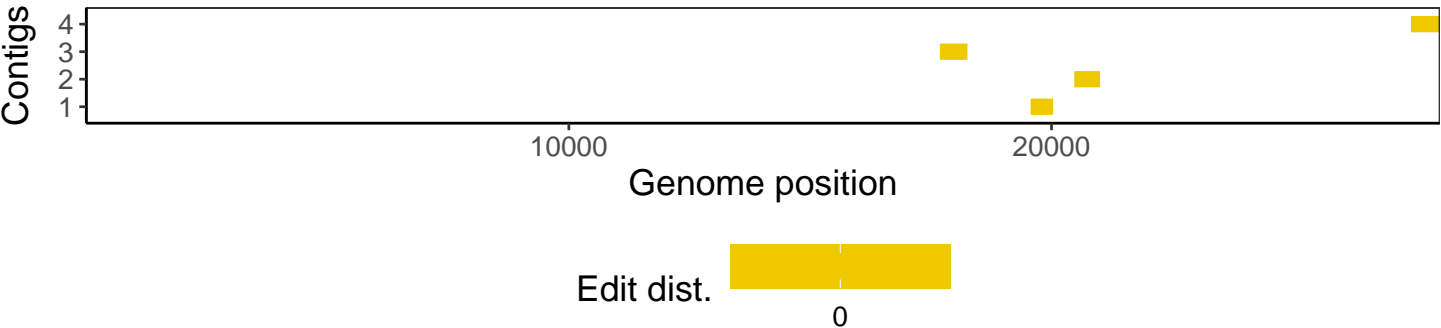
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



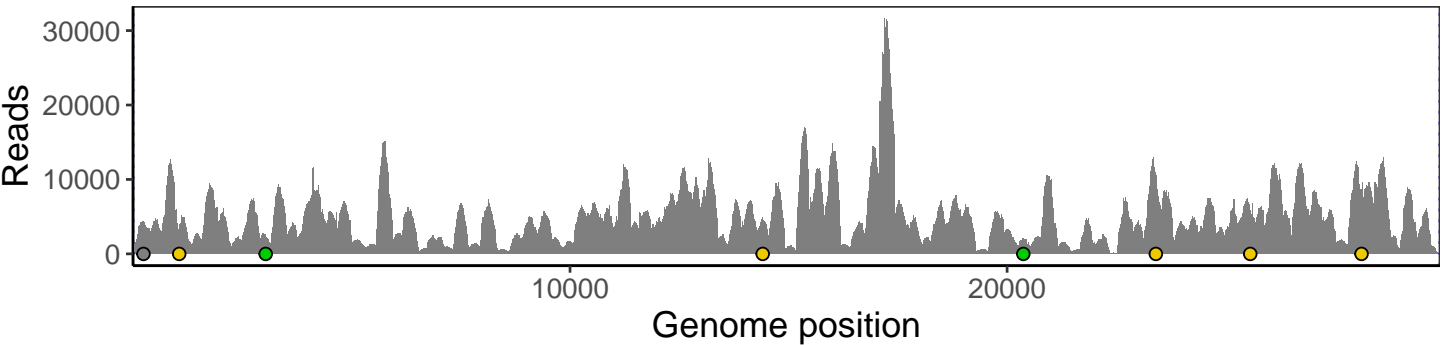
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



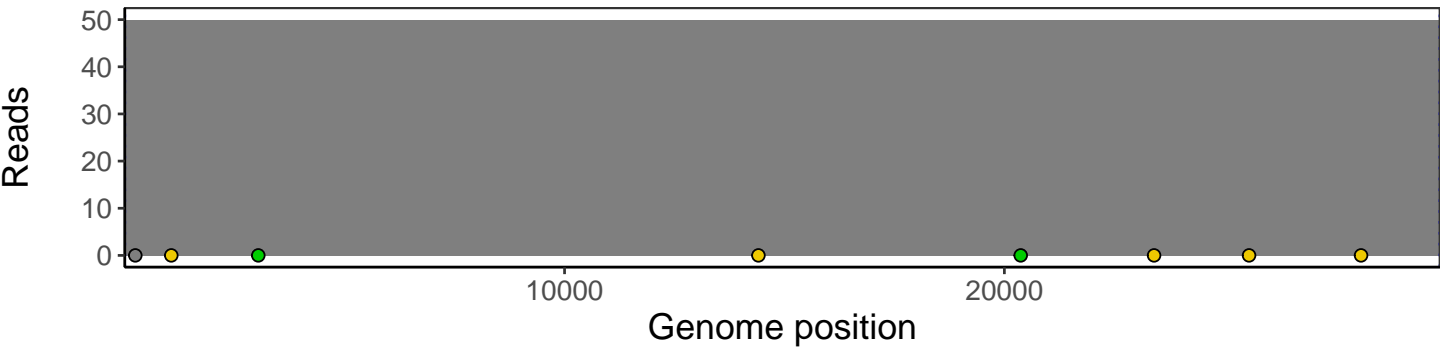
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



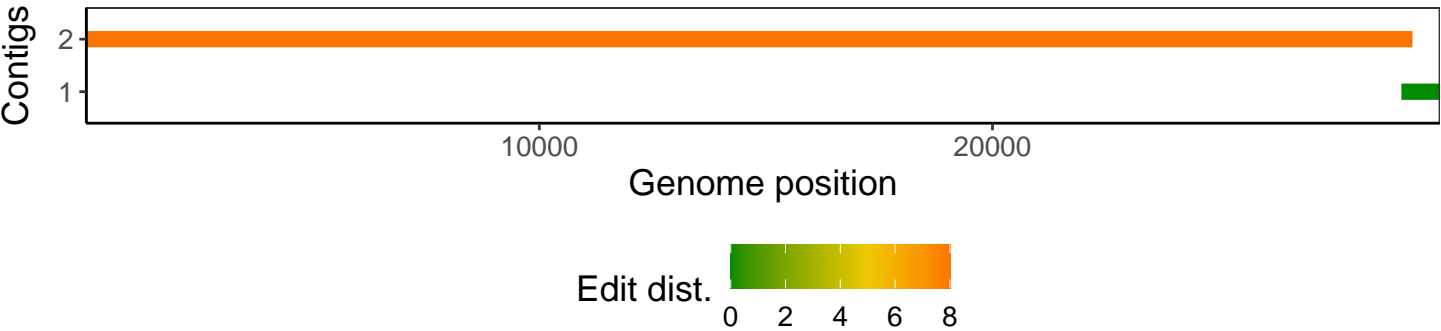
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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.



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