

COVID-19 subject 222-TCE

2020-09-14

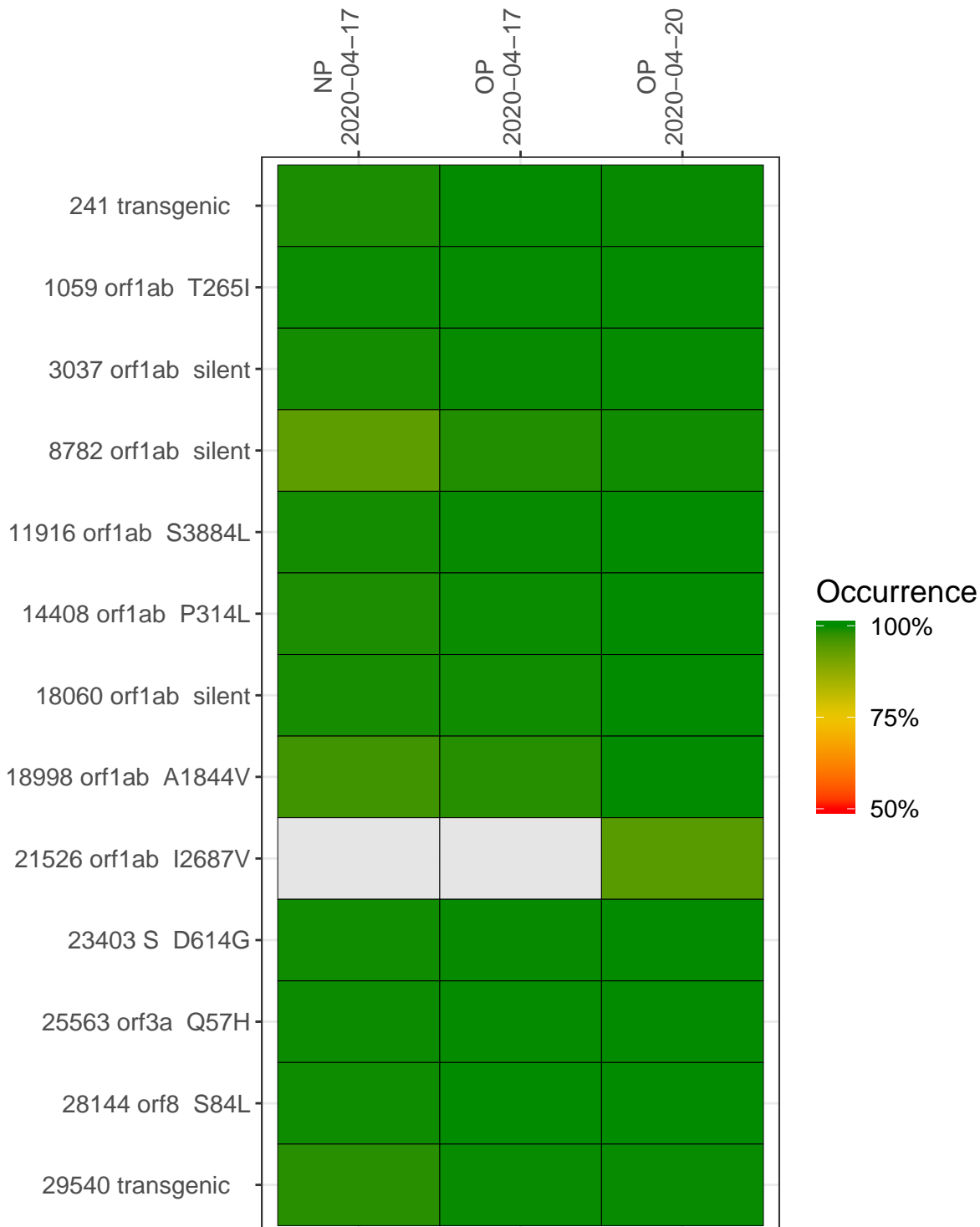
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. The code base for this analysis can be found ([here](#)).

Table 1. Sample summary.

Experiment	Type	Input genomes	Sample type	Sample date	Largest contig (KD)	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0181	composite	NA	OP	2020-04-17	29.34	99.9%	99.9%
VSP0182	composite	NA	OP	2020-04-20	29.88	99.9%	99.9%
VSP0180-1m	single experiment	NA	NP	2020-04-17	29.87	99.9%	99.7%
VSP0181-1m	single experiment	NA	OP	2020-04-17	30.04	99.9%	99.9%
VSP0181-2	single experiment	1.510e+11	OP	2020-04-17	29.34	99.9%	99.7%
VSP0182-1m	single experiment	NA	OP	2020-04-20	29.87	99.9%	99.9%
VSP0182-2	single experiment	1.325e+11	OP	2020-04-20	20.33	99.9%	99.4%

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 for 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 2020-04-17	OP 2020-04-17		OP 2020-04-20	
241 transgenic	753	704	576	741	729
1059 orf1ab T265I	3415	5162	562	2877	675
3037 orf1ab silent	1293	3513	716	1527	671
8782 orf1ab silent	106	283	489	281	219
11916 orf1ab S3884L	597	2366	258	946	118
14408 orf1ab P314L	918	3229	474	1252	120
18060 orf1ab silent	780	3162	412	1151	261
18998 orf1ab A1844V	87	248	394	324	137
21526 orf1ab I2687V	15	36	31	36	16
23403 S D614G	1517	1948	1034	1824	1065
25563 orf3a Q57H	2689	4793	822	2629	1069
28144 orf8 S84L	325	928	328	626	214
29540 transgenic	131	757	374	265	316
	VSP0180-1m	VSP0181-1m	VSP0181-2	VSP0182-1m	VSP0182-2

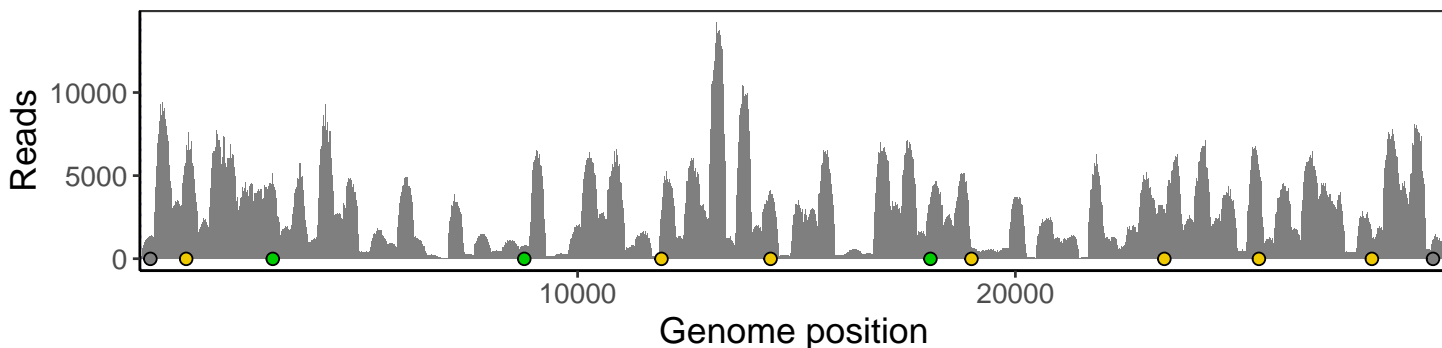
Base change



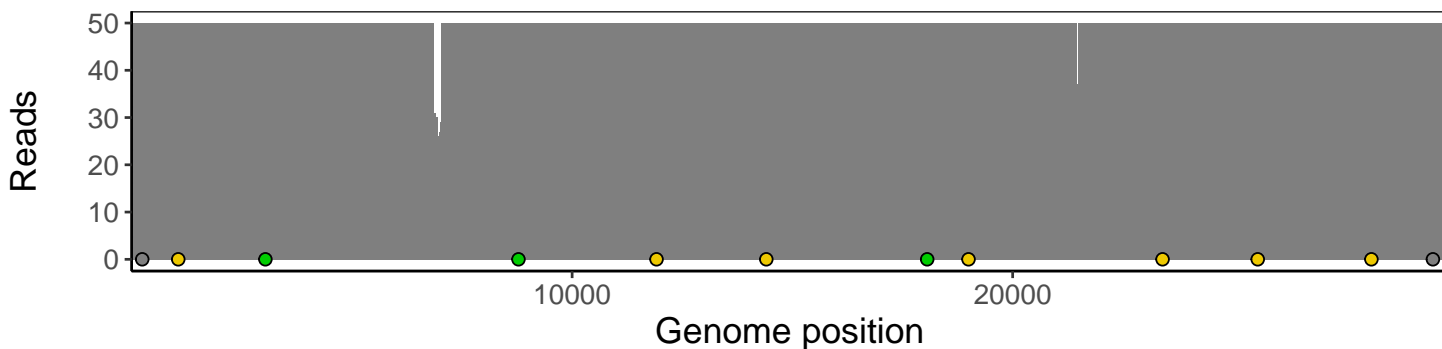
Analyses of individual experiments and composite results.

VSP0181 | 2020-04-17 | OP | 2 | composite result

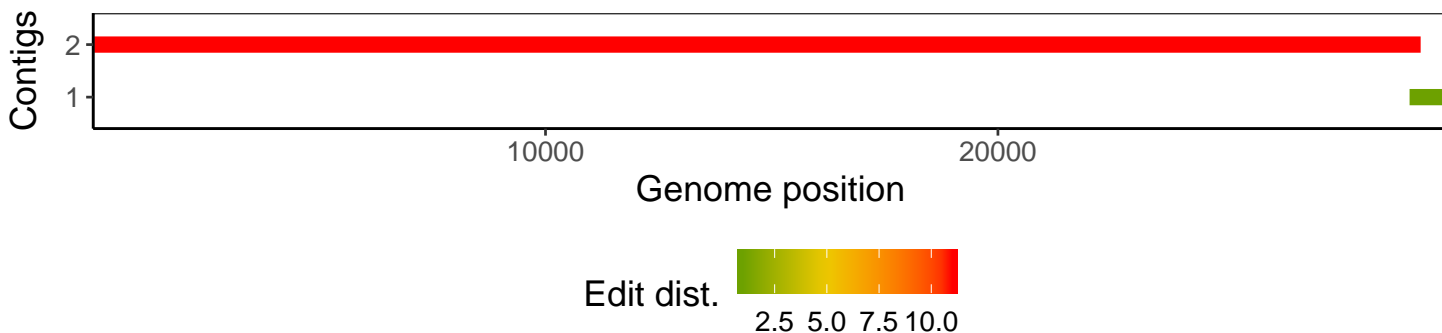
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according to variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



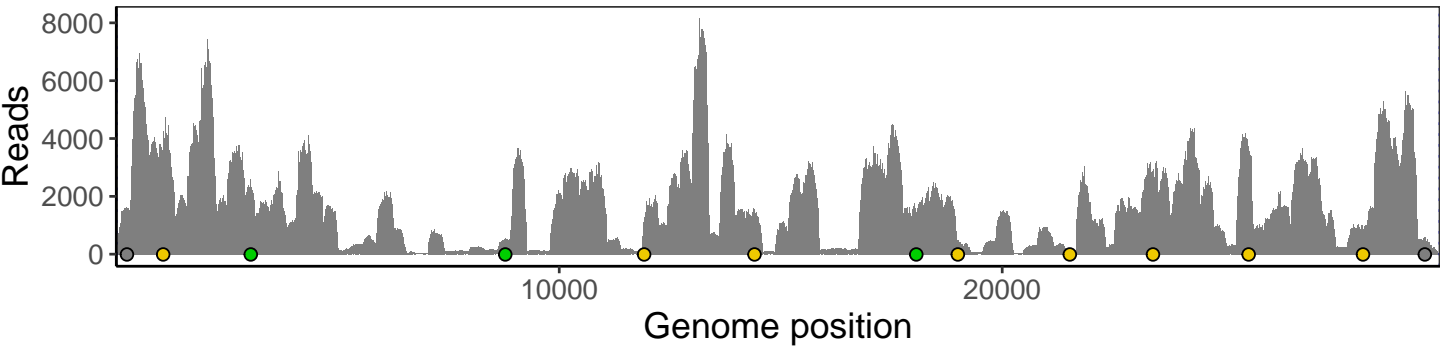
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



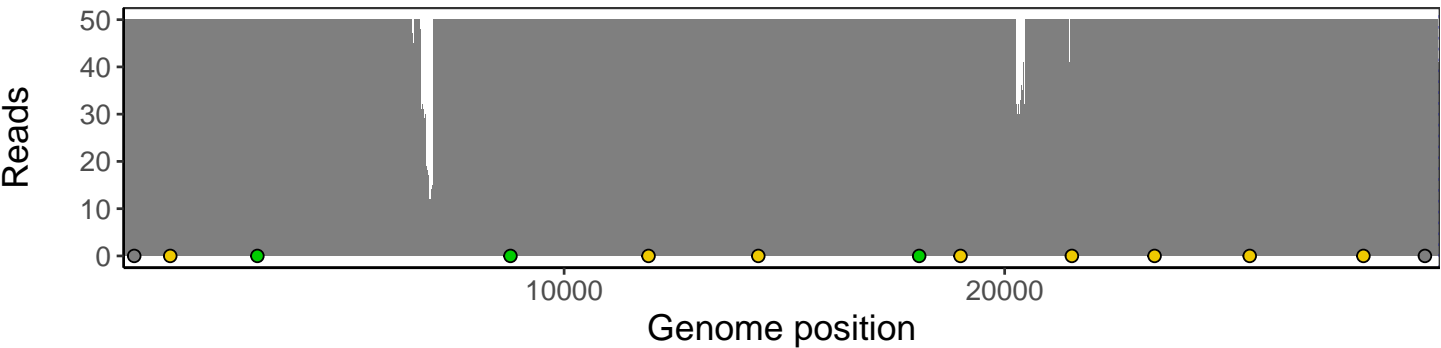
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



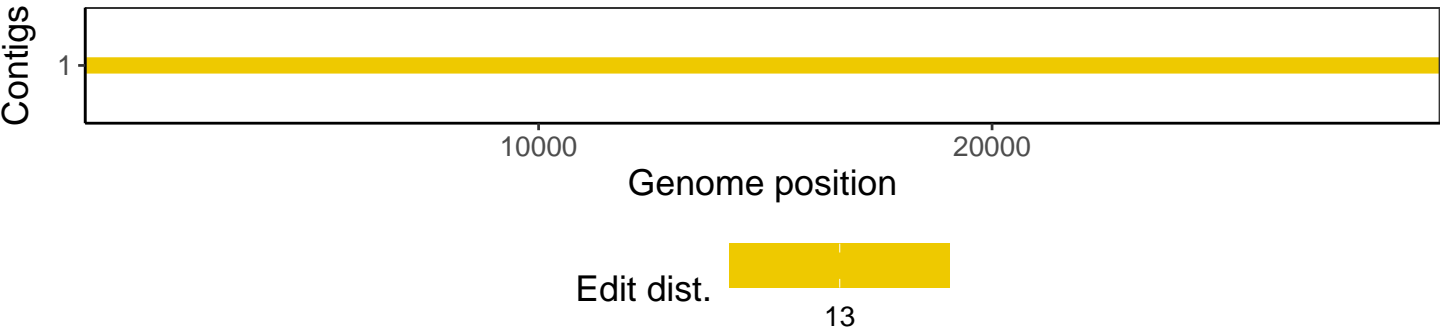
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



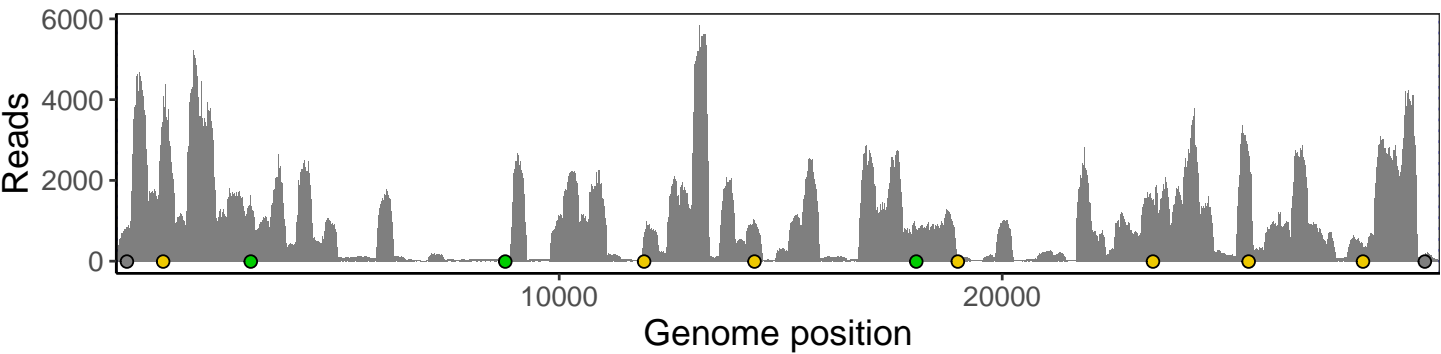
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



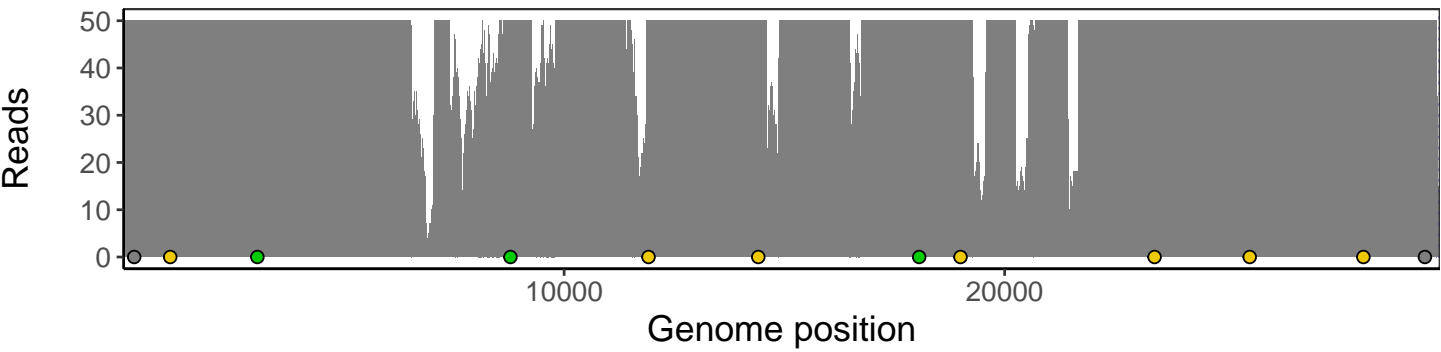
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



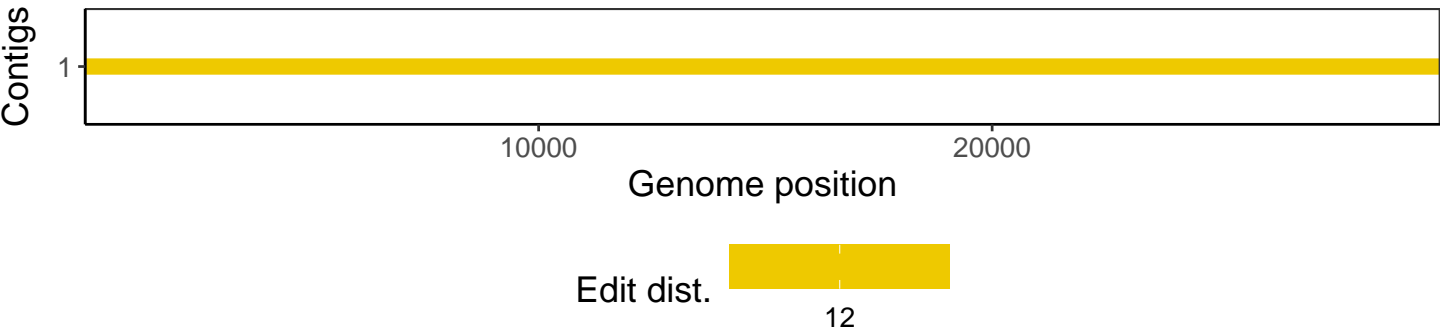
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



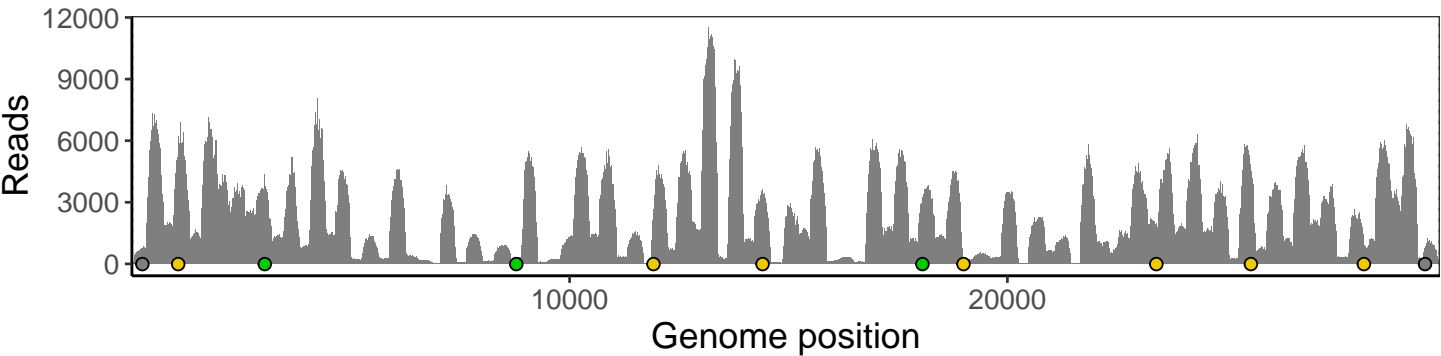
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



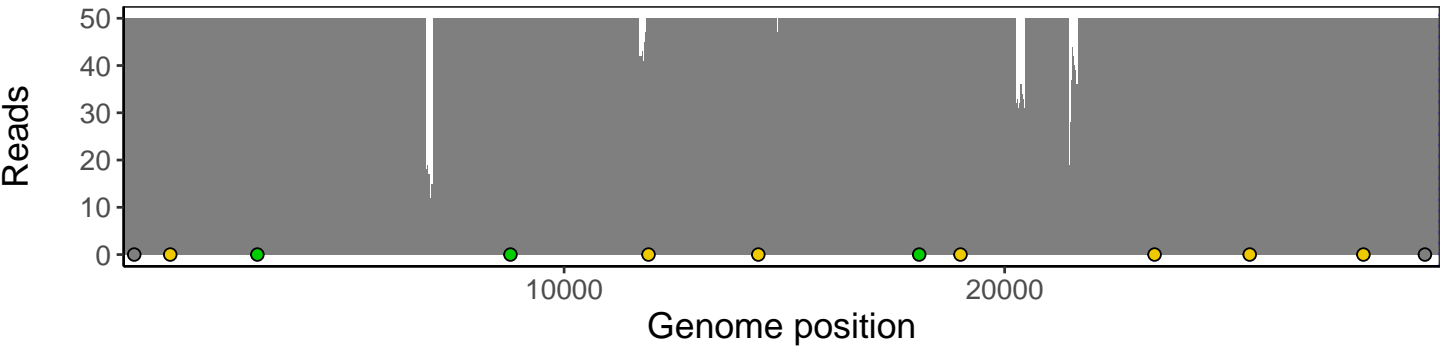
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



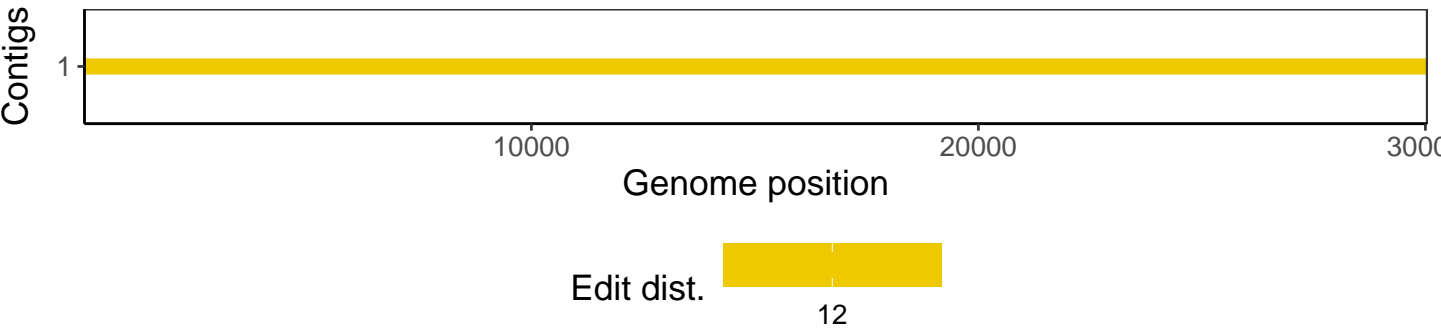
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



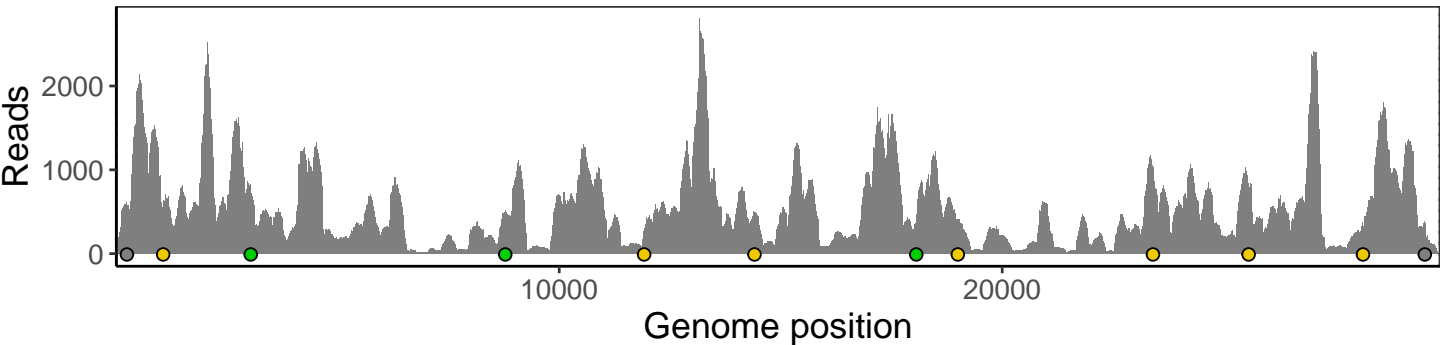
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



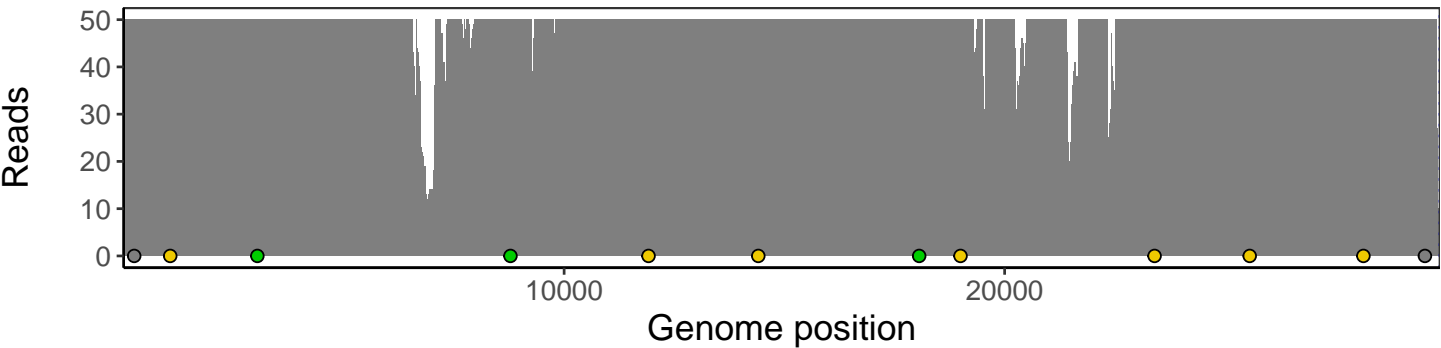
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



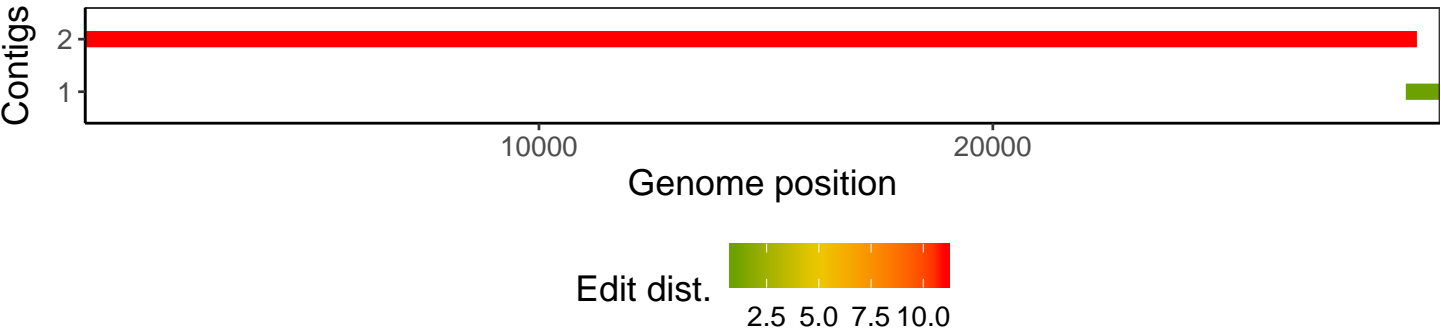
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



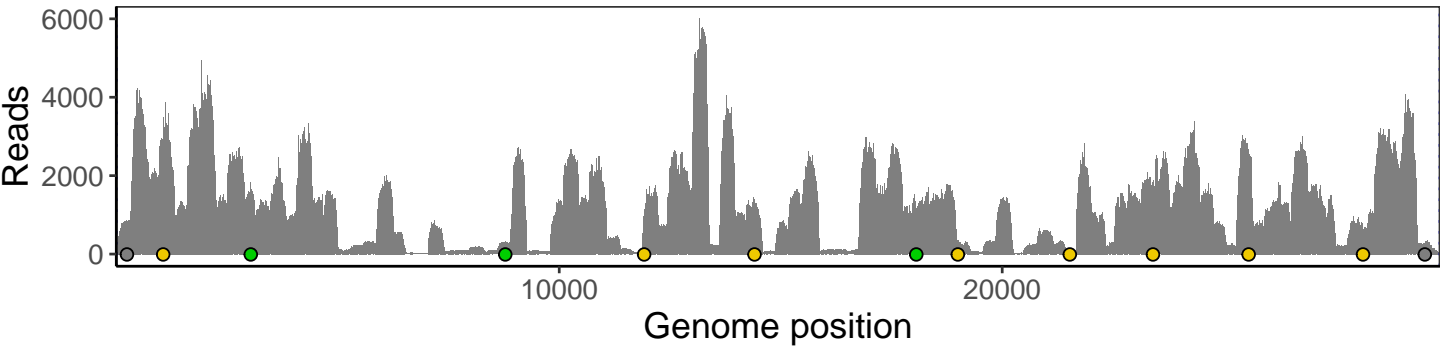
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



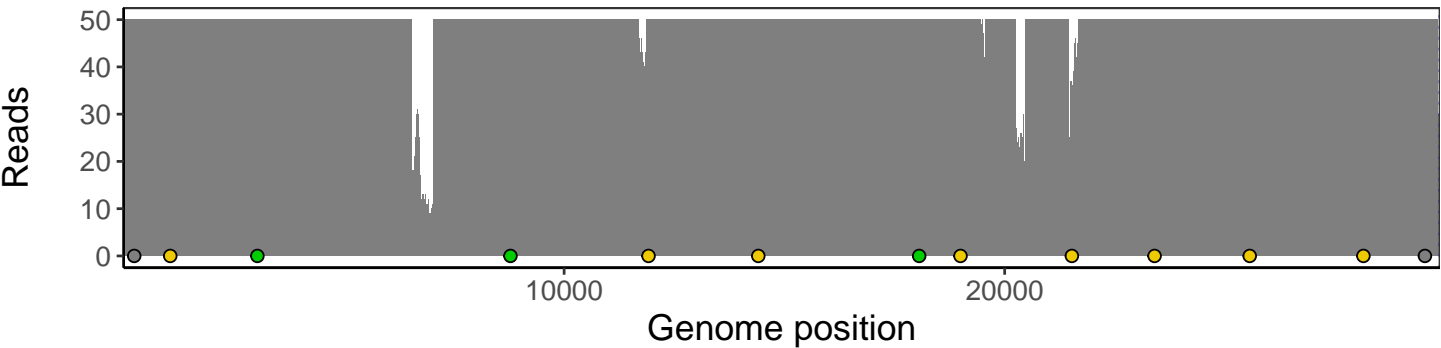
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



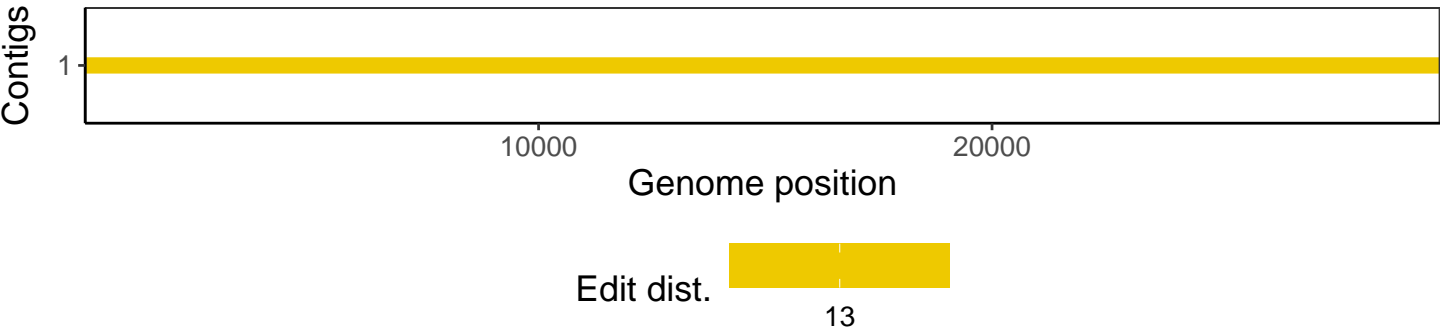
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



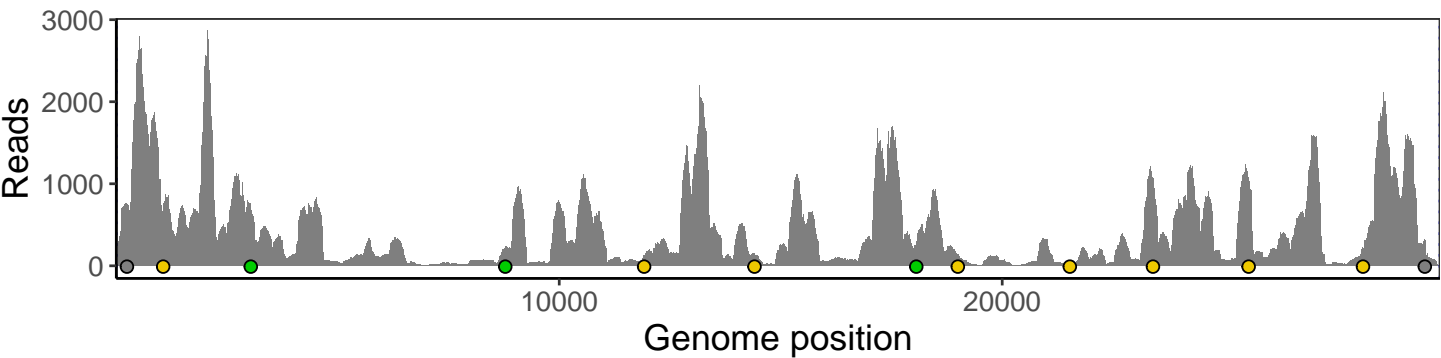
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



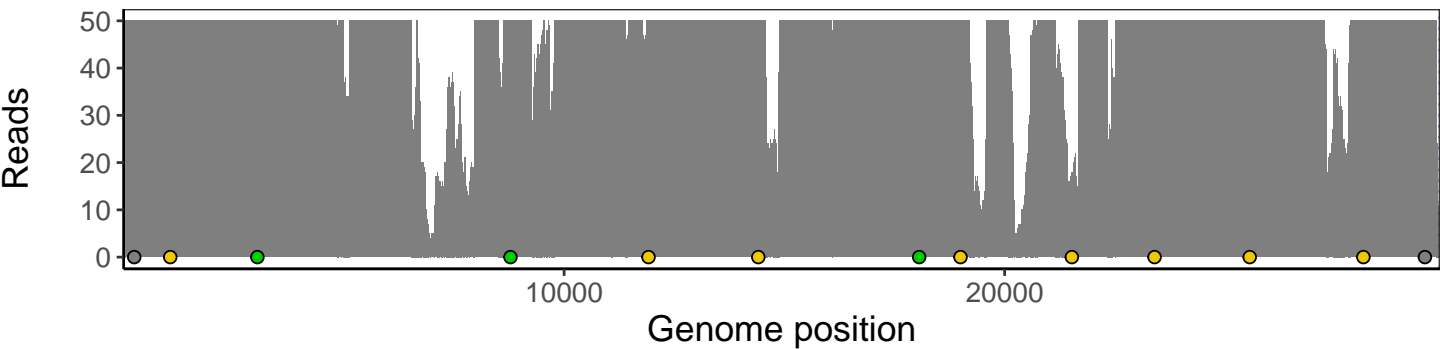
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



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.

