# COVID-19 subject 269

2021-05-05

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)
VSP0177	composite	NA	NP-OP	2020-05-27	4.03	NA	90.1%	79.9%
VSP0166-1	single experiment	NA	NP-OP	2020-05-22	29.84	B.1.369	99.9%	99.7%
VSP0176-1	single experiment	166000.0	ETA	2020-05-27	29.82	B.1.369	99.9%	99.8%
VSP0177-1	single experiment	59.7	NP-OP	2020-05-27	1.03	NA	72.8%	38.6%
VSP0177-2	single experiment	NA	NP-OP	2020-05-27	1.01	NA	31.8%	27.6%
VSP0177-3	single experiment	NA	NP-OP	2020-05-27	1.02	NA	40.9%	38.3%
VSP0177-4	single experiment	NA	NP-OP	2020-05-27	0.93	NA	38.7%	36.3%
VSP0200-1	single experiment	NA	ETA	2020-06-08	5.92	NA	84.7%	82.6%
VSP0202-1	single experiment	16300.0	ETA	2020-06-12	29.83	B.1.369	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.
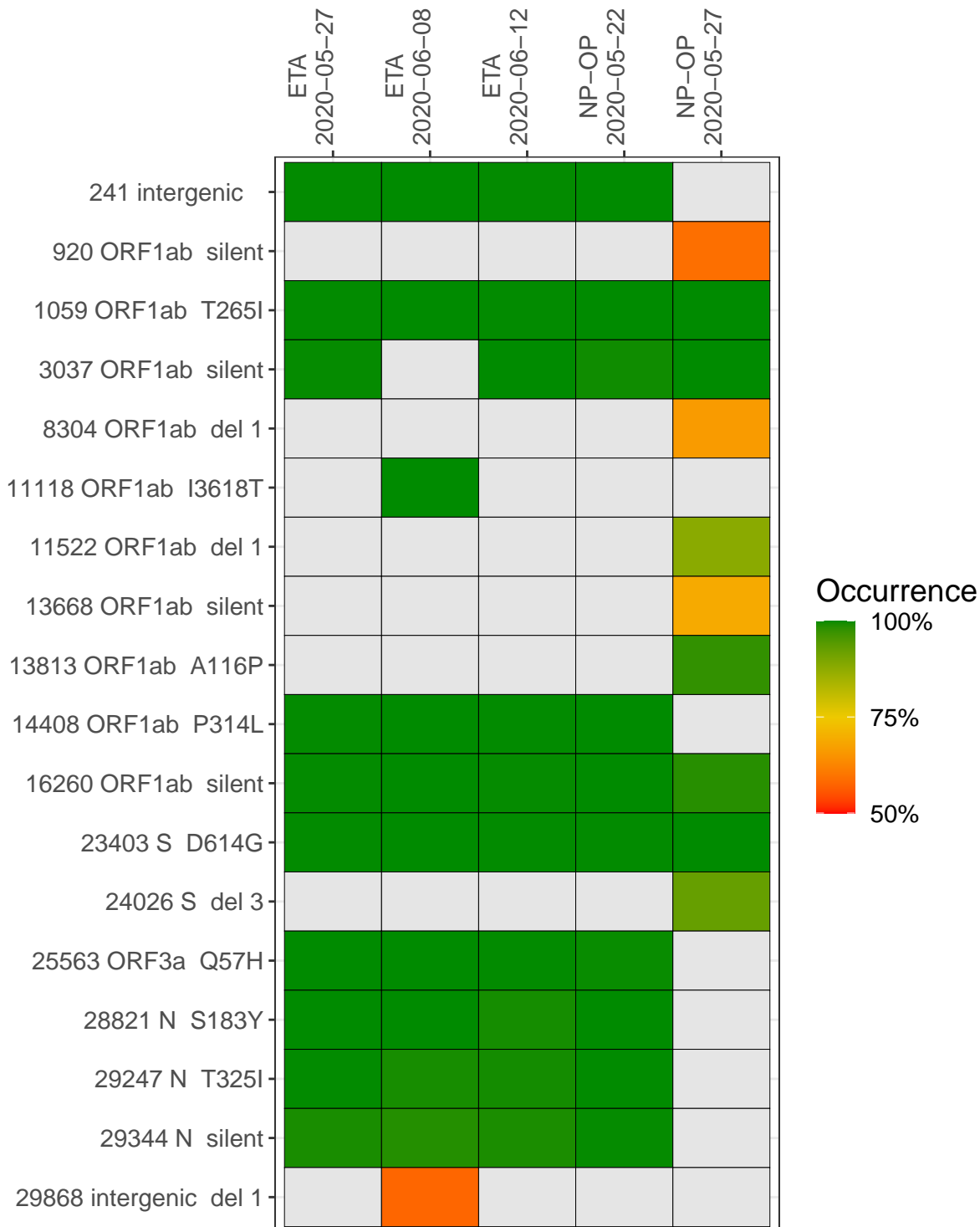
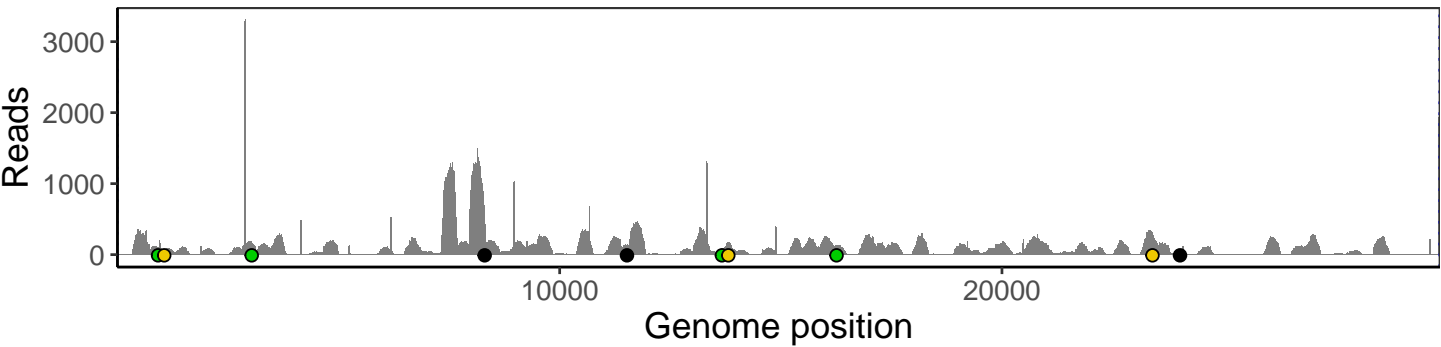|	ETA 2020-05-2	ETA 2020-06-0	ETA 2020-06-1	NP-OP 2020-05-2	NP-OP 2020-05-27			
241 intergenic	1809	421	3078	1755	4	0	0	0
920 ORF1ab silent	2399	164	2037	1791	3	0	102	0
1059 ORF1ab T265I	1968	91	2625	1147	8	47	13	0
3037 ORF1ab silent	1826	0	1837	2722	6	62	26	72
8304 ORF1ab del 1	890	180	3673	699	305	36	20	81
11118 ORF1ab I3618T	1848	276	5694	2041	8	28	99	0
11522 ORF1ab del 1	2028	1099	5704	4050	16	0	121	0
13668 ORF1ab silent	1239	246	3056	2315	10	58	14	2
13813 ORF1ab A116P	2752	771	6084	4094	4	172	0	0
14408 ORF1ab P314L	2514	448	4115	1863	4	0	0	0
16260 ORF1ab silent	2383	70	3398	2972	2	0	0	130
23403 S D614G	7570	844	9833	3918	10	157	119	0
24026 S del 3	929	0	1791	1469	2	0	37	0
25563 ORF3a Q57H	1619	144	5108	2099	9	0	0	0
28821 N S183Y	1331	48	1825	1599	0	0	0	0
29247 N T325I	2254	307	3391	4033	0	0	0	0
29344 N silent	1693	148	2586	2605	2	0	0	0
29868 intergenic del 1	7	69	10	0	12	7	7	4
	VSP0176-1	VSP0200-1	VSP0202-1	VSP0166-1	VSP0177-1	VSP0177-2	VSP0177-3	VSP0177-4

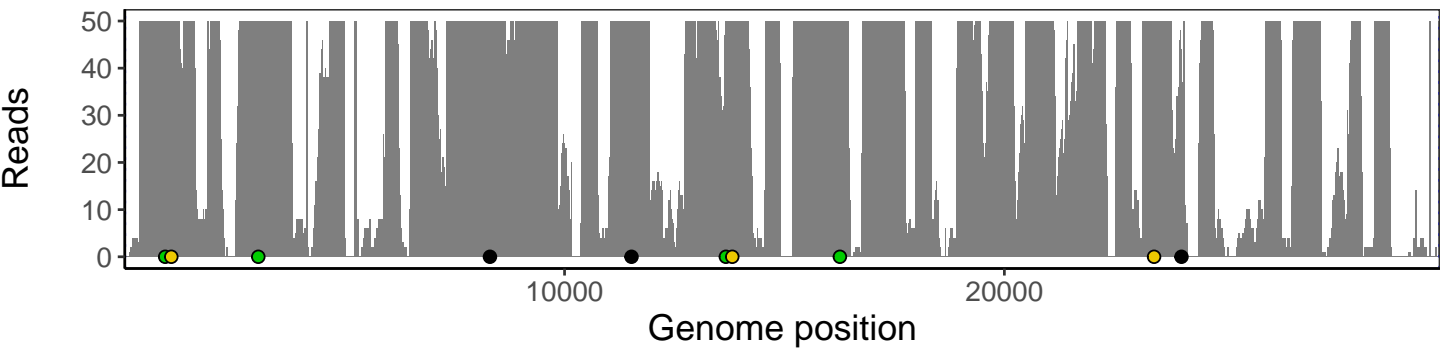Base change
Expected
A
T
C
G
N
Ins/Del
No data

Analyses of individual experiments and composite results

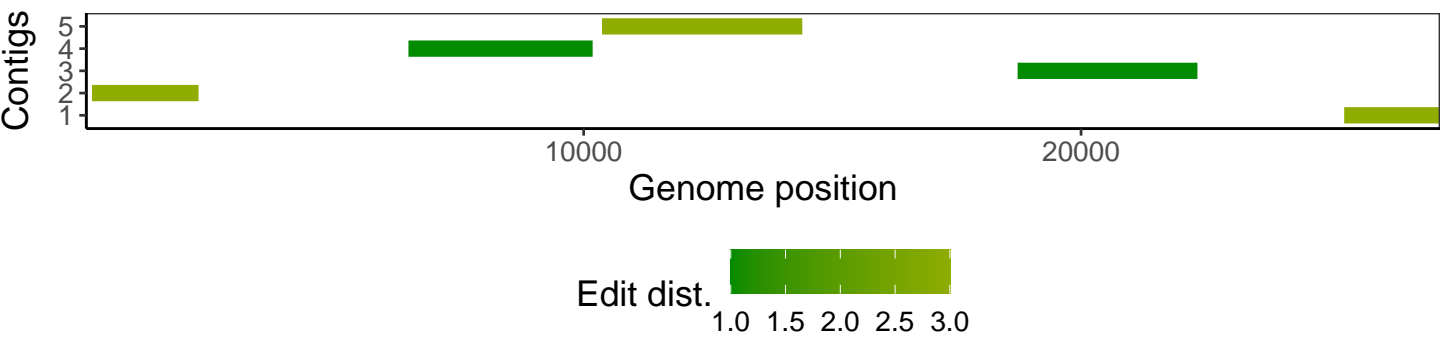**VSP0177 | 2020-05-27 | NP-OP | 269no-q | composite result**

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.
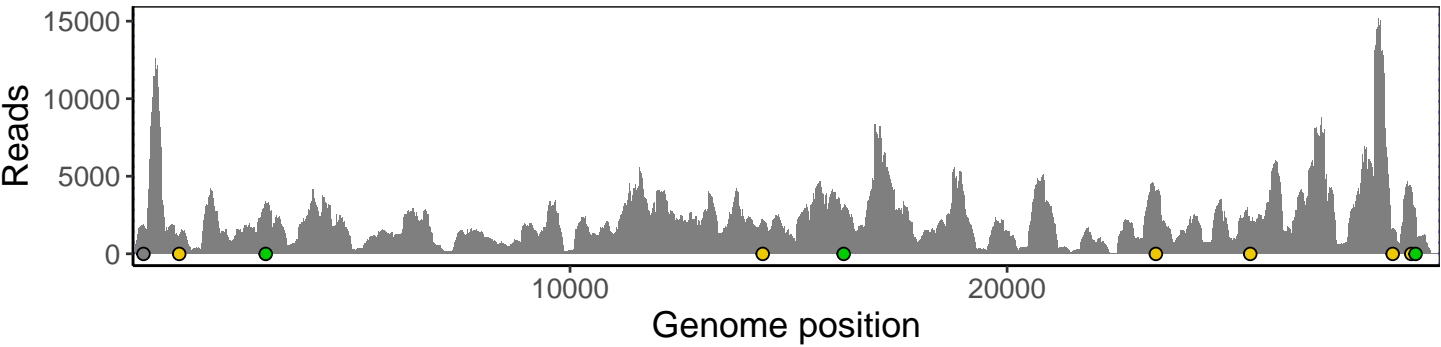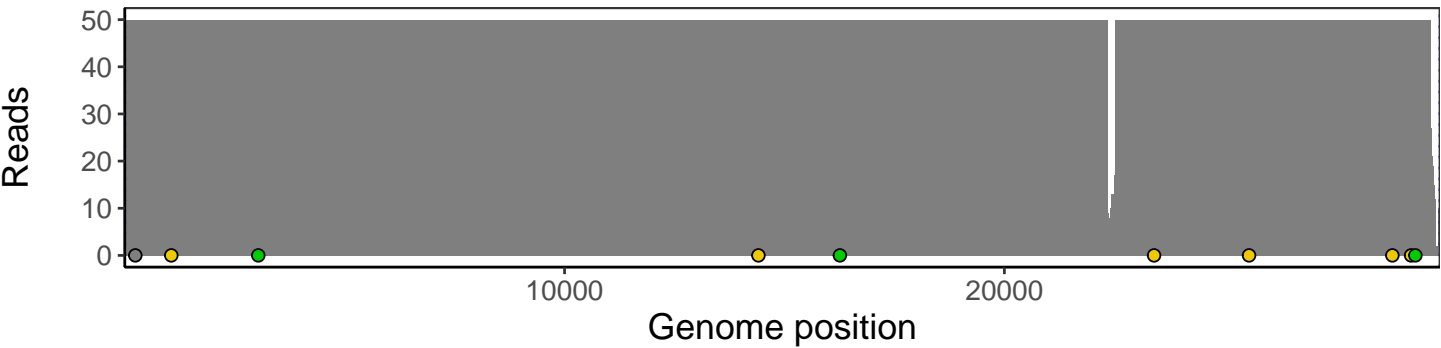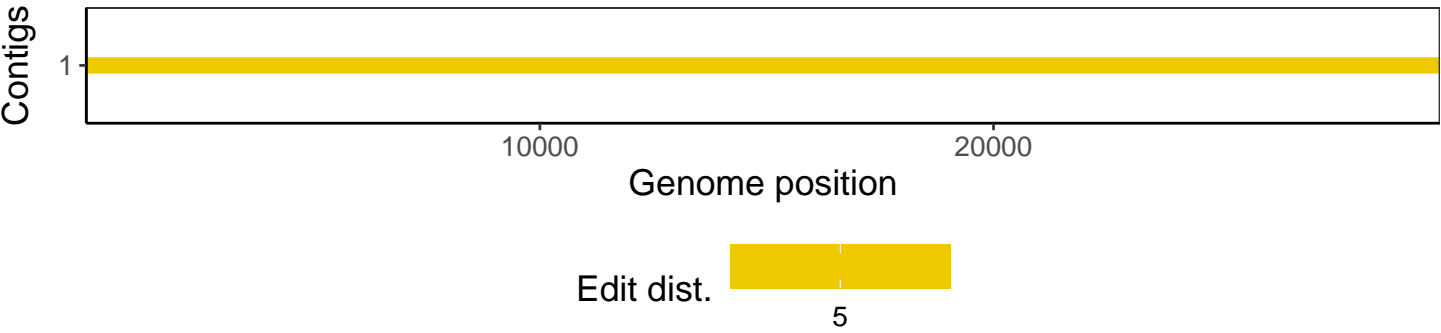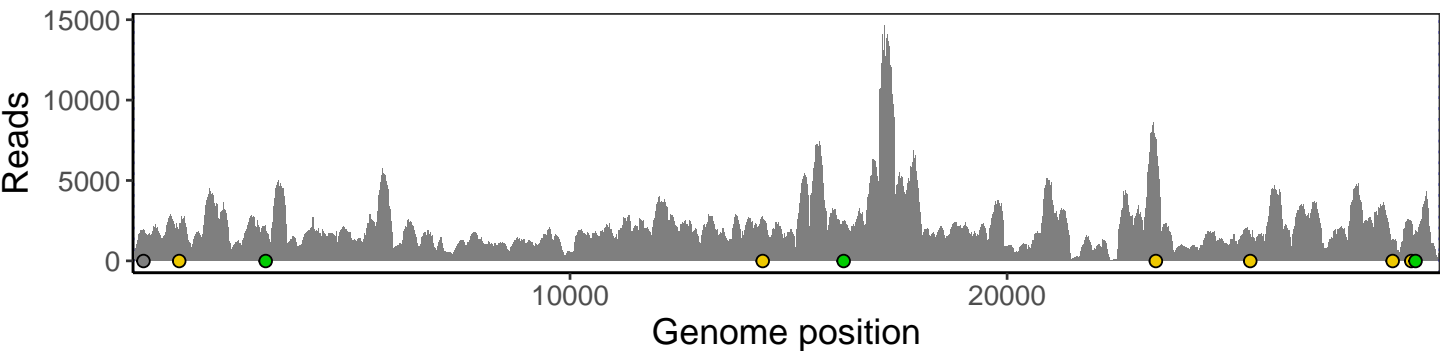

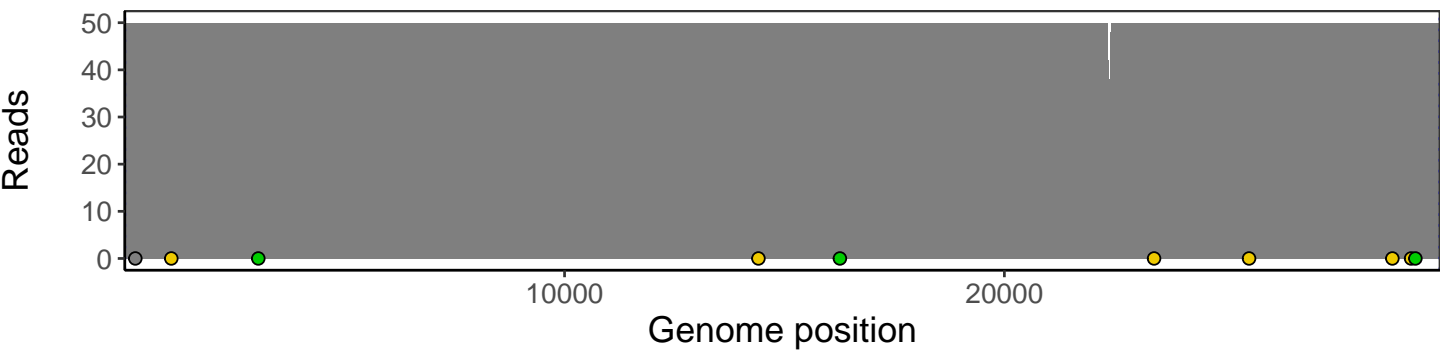Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

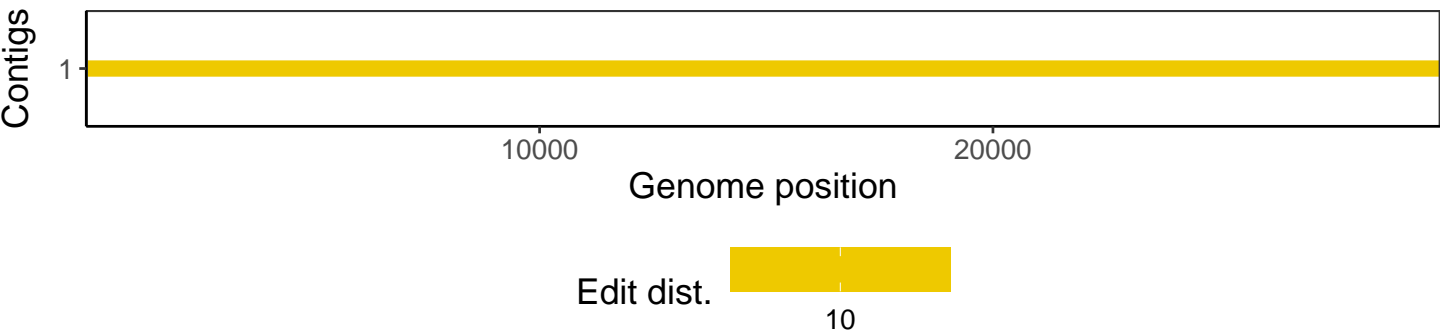**VSP0166-1 | 2020-05-22 | NP-OP | 269no-q | genomes | single experiment**

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.
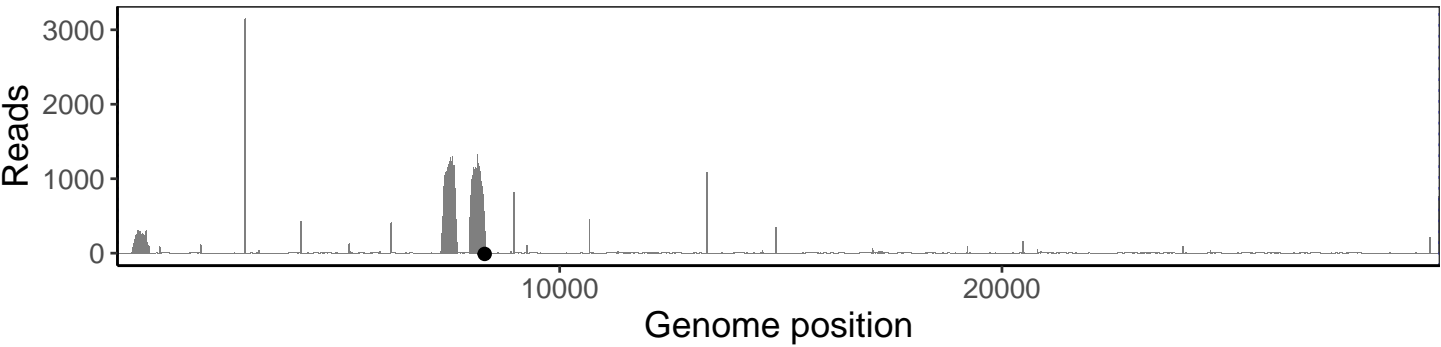

Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
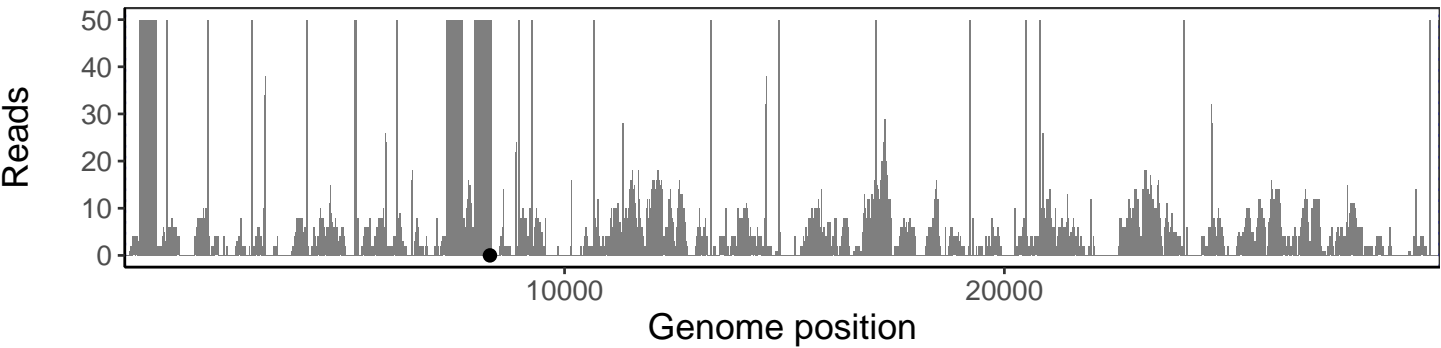

The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

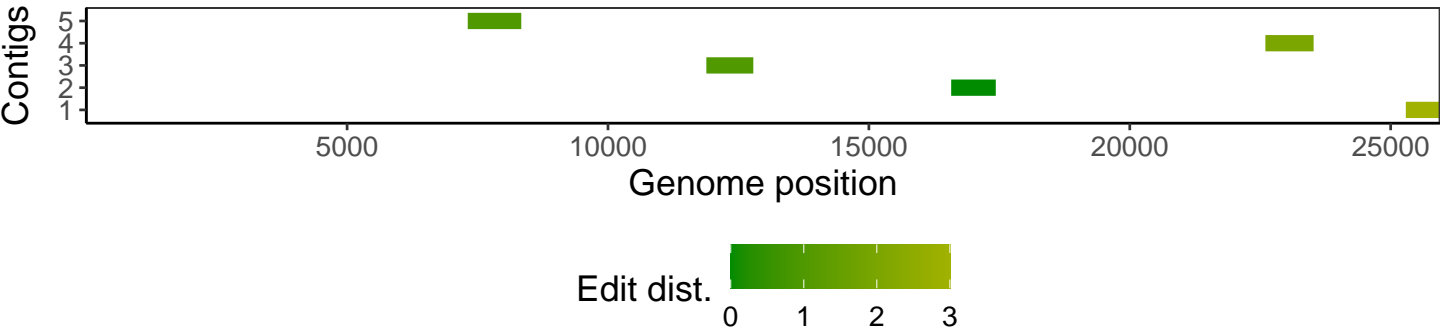**VSP0176-1 | 2020-05-27 | ETA | 269e-q | 166000 genomes | single experiment**

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.
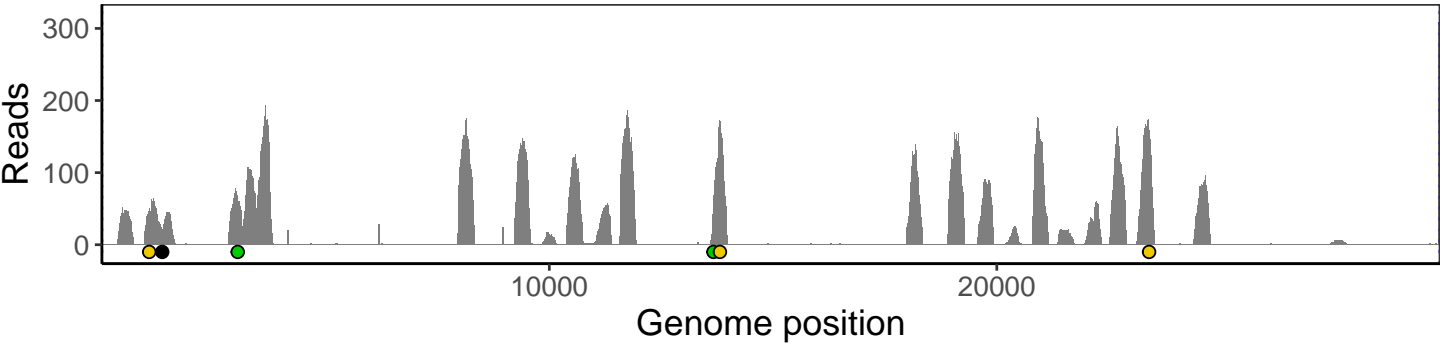

Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

VSP0177-1 | 2020-05-27 | NP-OP | 269no-q | 59.7 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.

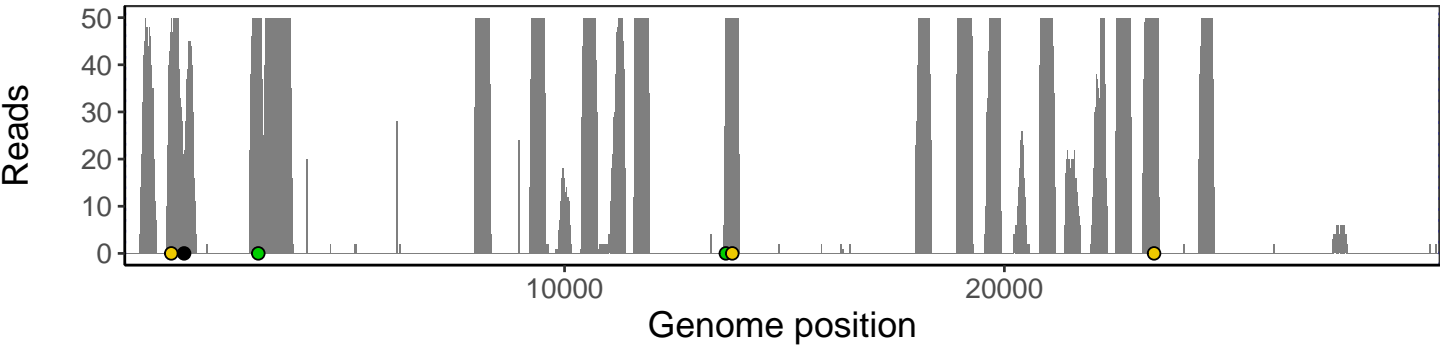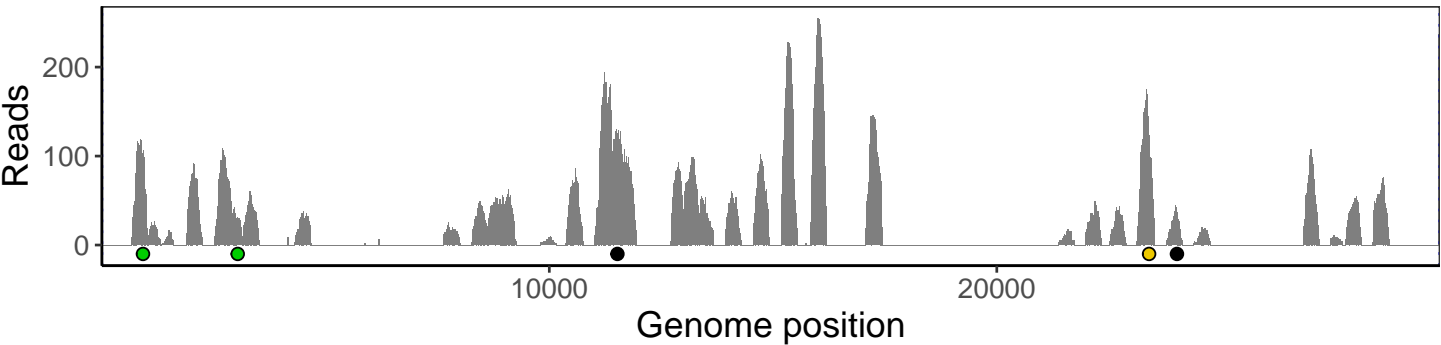Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

VSP0177-2 | 2020-05-27 | NP-OP | 269no-q | 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.
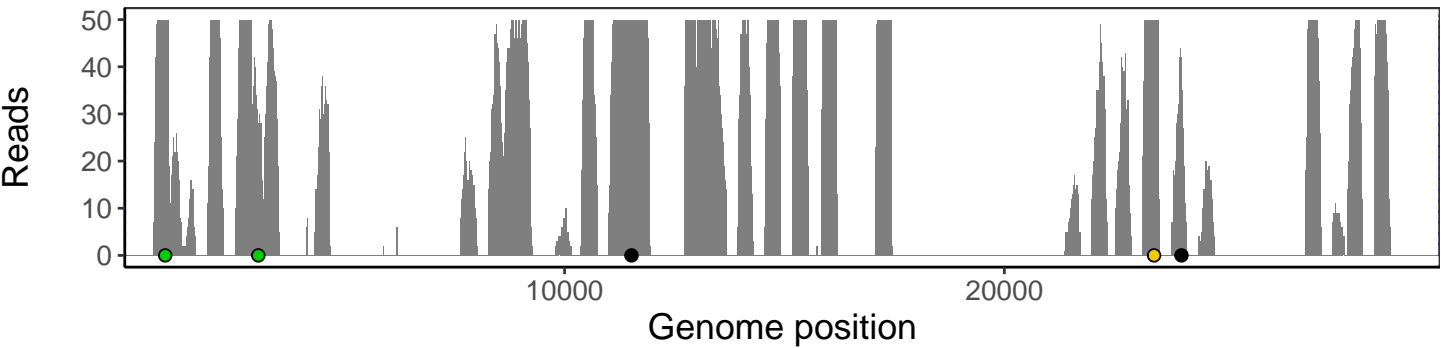

The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

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.
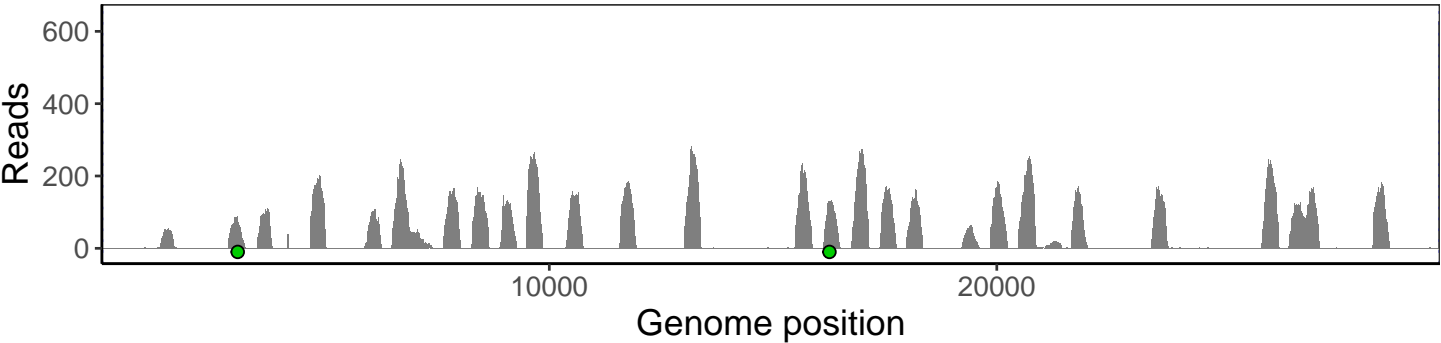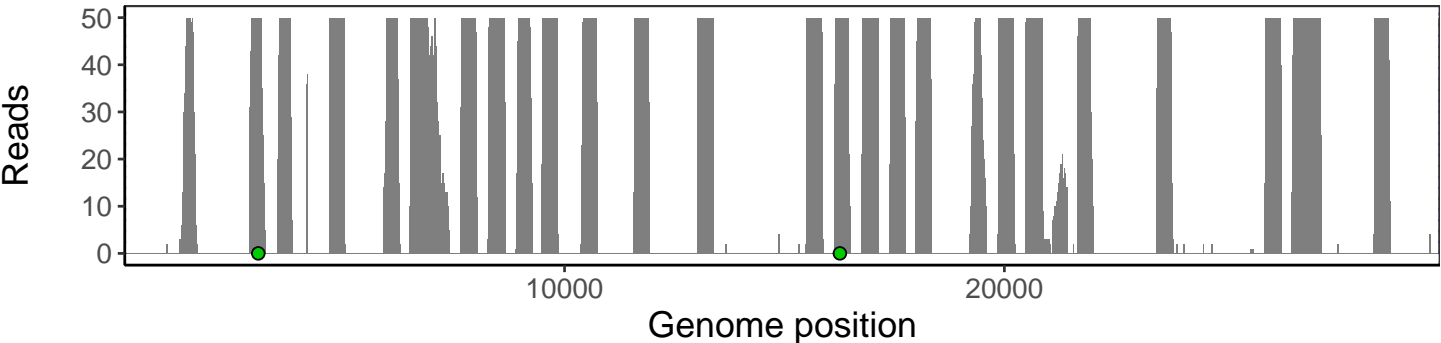

Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

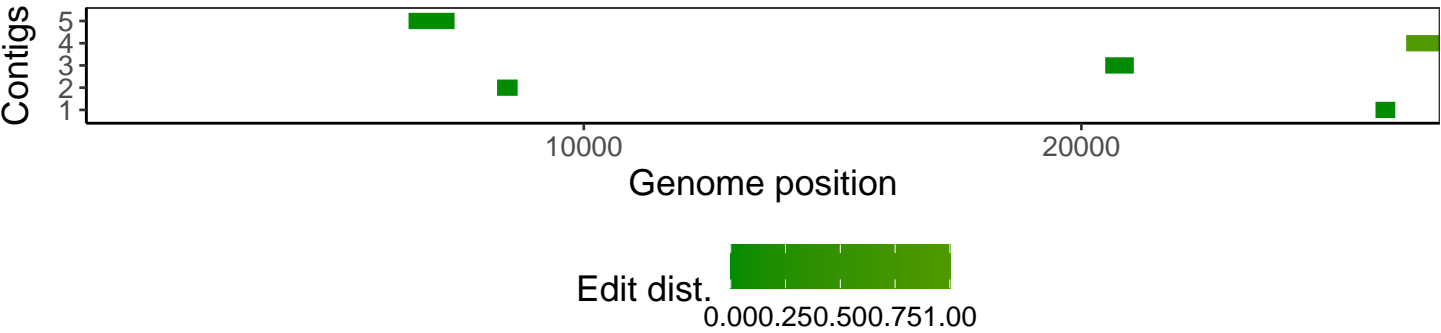**VSP0177-4 | 2020-05-27 | NP-OP | 269no-q | genomes | single experiment**

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.
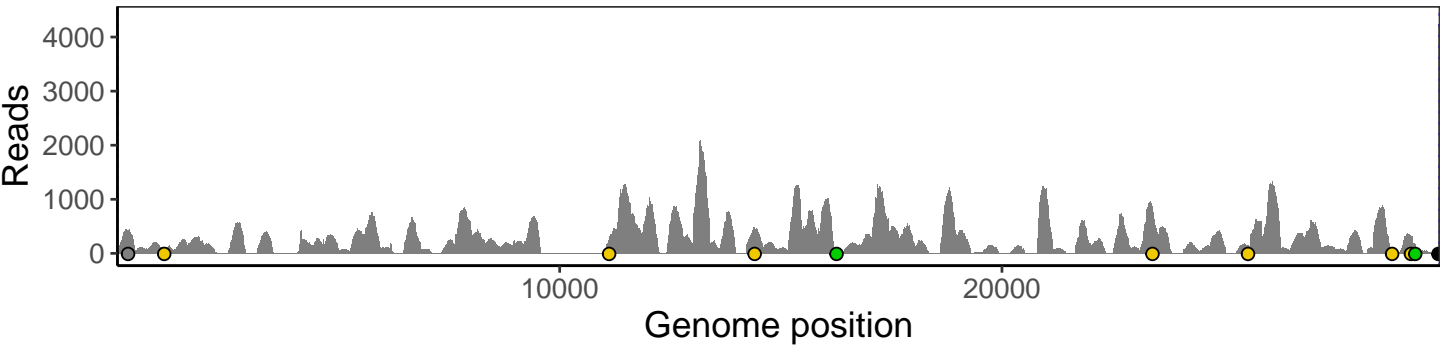

Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

VSP0200-1 | 2020-06-08 | ETA | 269e-q | 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.

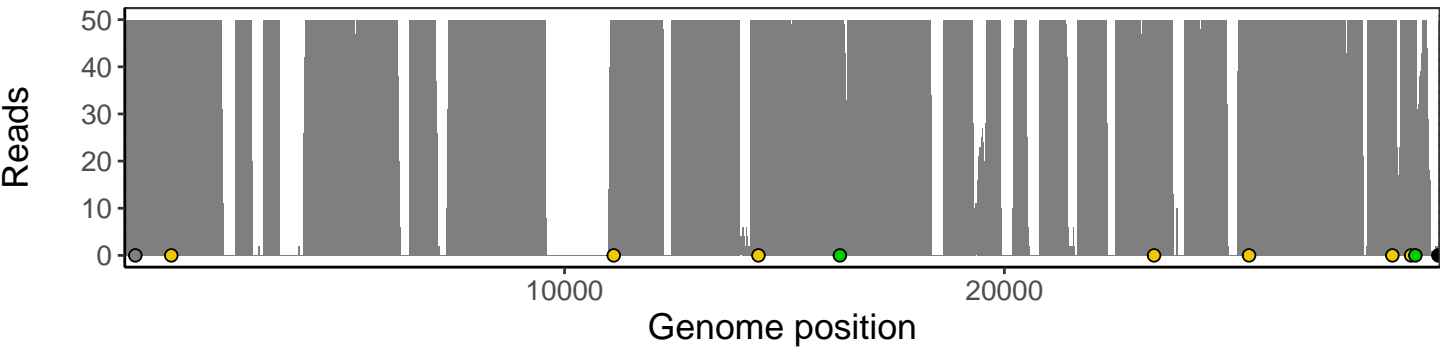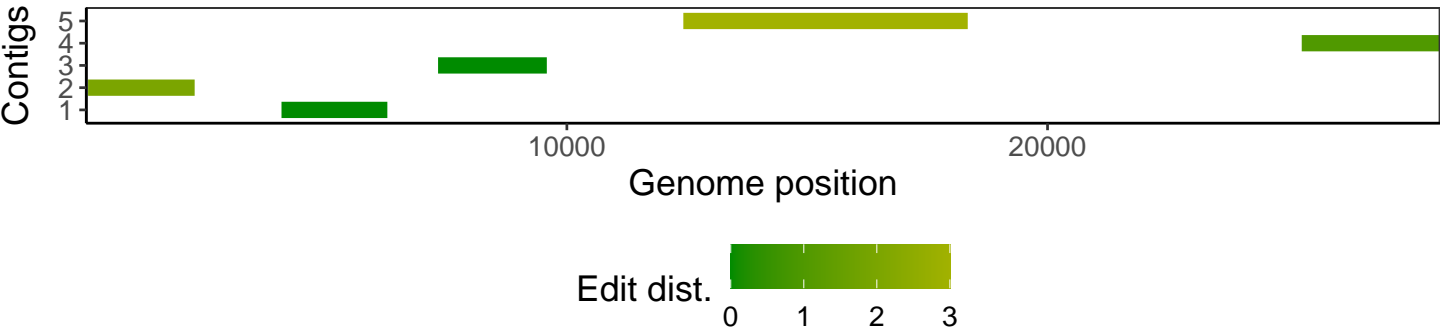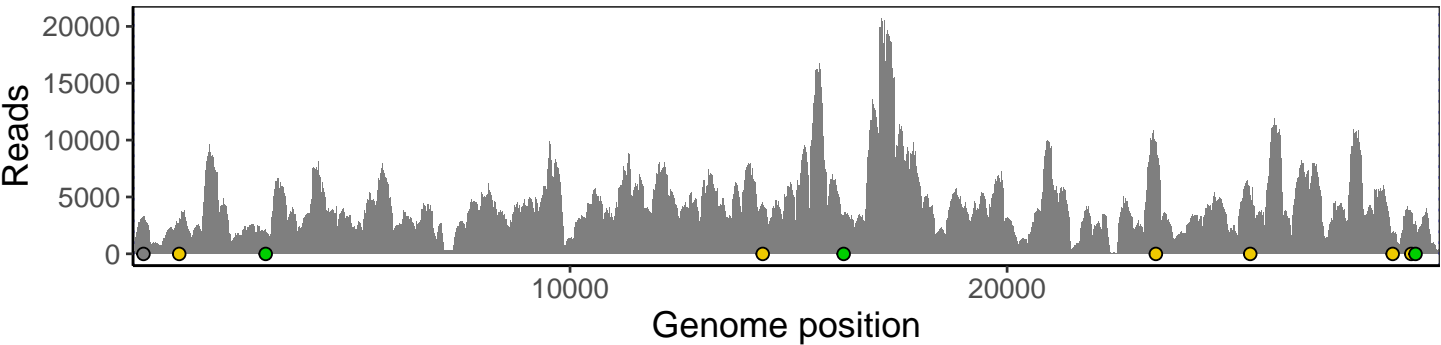Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

VSP0202-1 | 2020-06-12 | ETA | 269e-q | 16300 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.
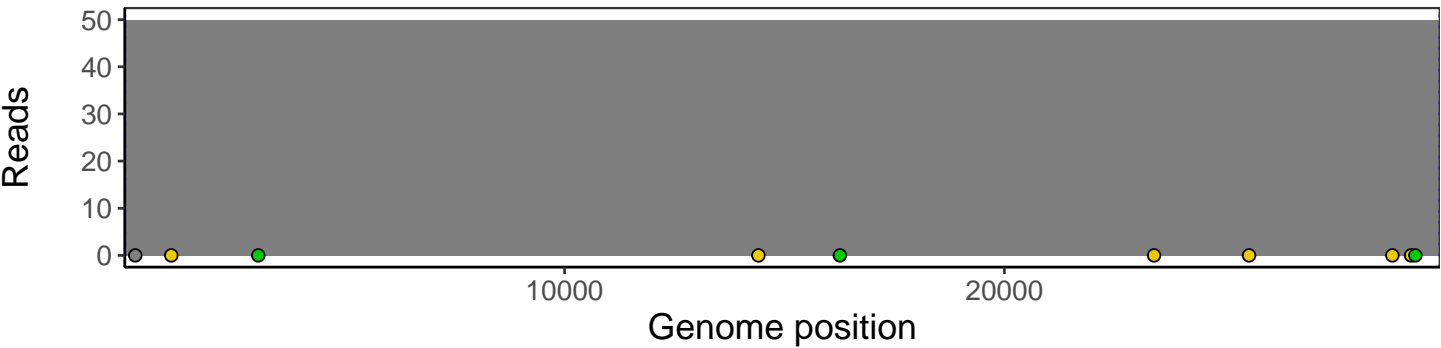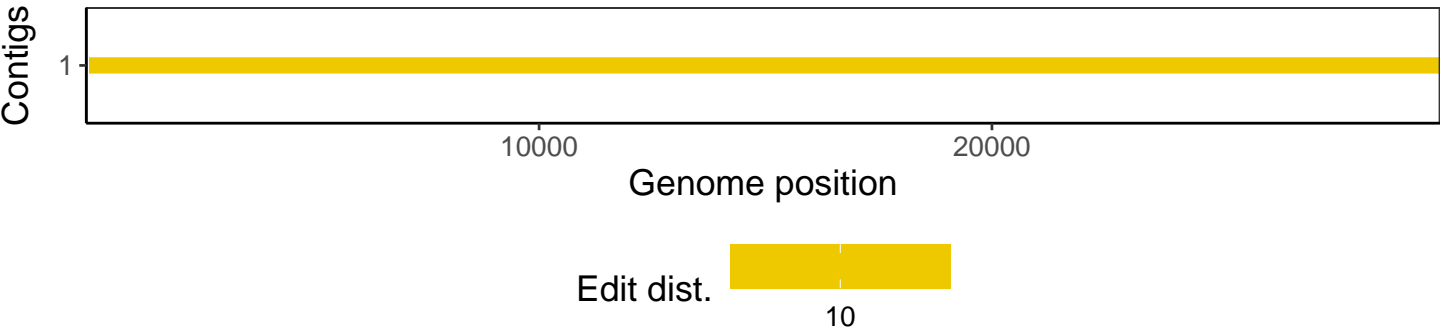

The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

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	2.3.8
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