# COVID-19 subject 251

2021-01-27

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)
VSP0065	composite	NA	NP-OP	2020-05-04	16.53	B.1	99.9%	99.8%
VSP0089	composite	NA	NP-OP	2020-05-06	30.00	B.1	99.9%	99.7%
VSP0065-1	single experiment	7550000	NP-OP	2020-05-04	16.47	B.1	99.9%	99.8%
VSP0065-2	single experiment	7550000	NP-OP	2020-05-04	29.85	B.1	99.8%	99.6%
VSP0088-1	single experiment	255500	ETA	2020-05-06	29.82	B.1	99.8%	99.8%
VSP0089-1	single experiment	570000	NP-OP	2020-05-06	29.91	B.1	99.9%	99.7%
VSP0089-2	single experiment	570000	NP-OP	2020-05-06	29.45	B.1	99.4%	99.1%
VSP0318-1	single experiment	890000	NP-OP	2020-05-06	24.51	B.1	99.8%	99.7%

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.
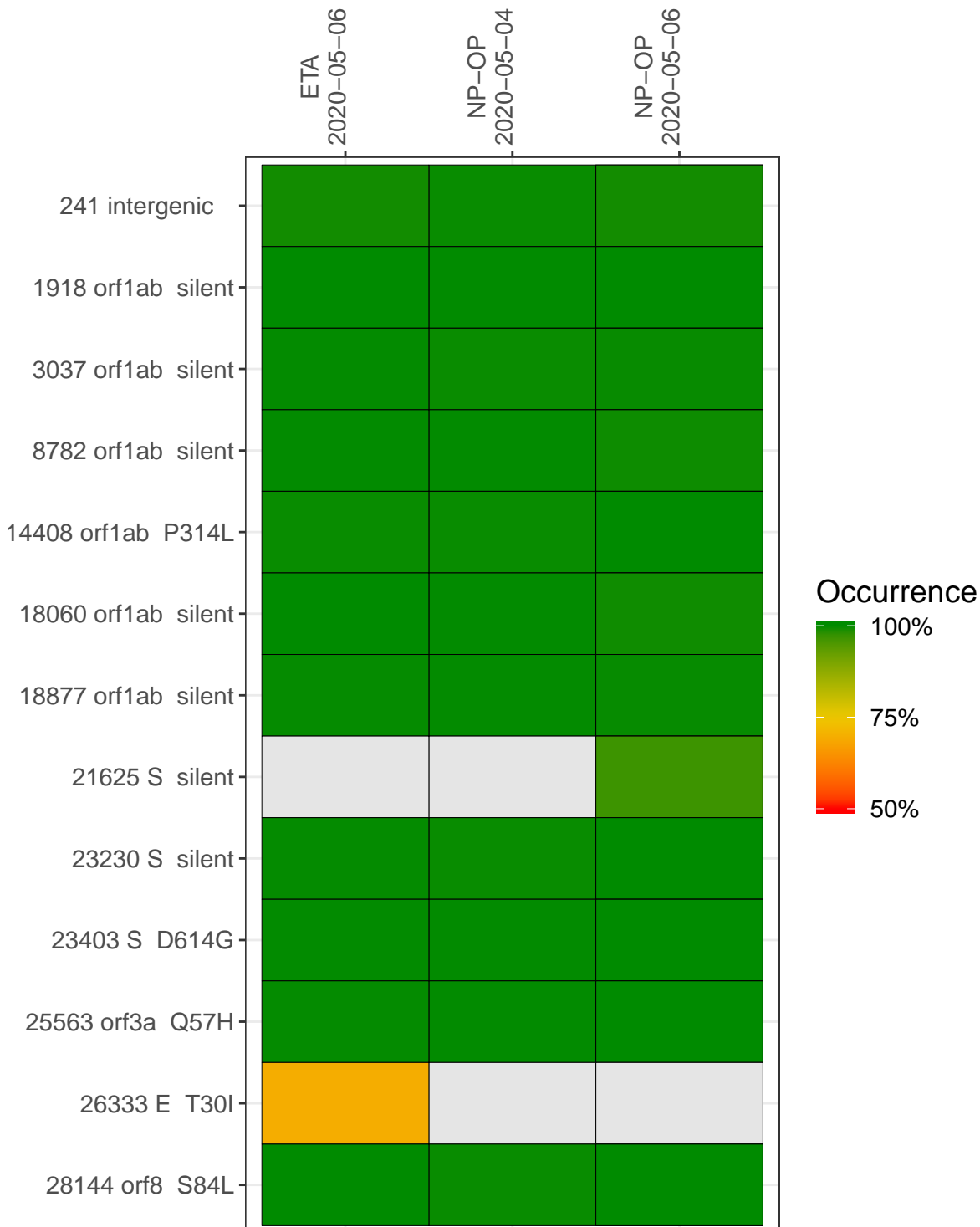
	ETA 2020-05-06	NP-OP 2020-05-04		NP-OP 2020-05-06		
241 intergenic	2812	2551	9904	4496	1124	1174
1918 orf1ab silent	2850	2440	8169	3070	2443	1004
3037 orf1ab silent	2431	5091	3037	3634	544	1756
8782 orf1ab silent	2242	4518	459	3782	205	623
14408 orf1ab P314L	5120	4016	127	7513	368	403
18060 orf1ab silent	2382	4745	927	3456	324	545
18877 orf1ab silent	6407	8094	333	8408	1475	1476
21625 S silent	518	2969	32	1039	67	158
23230 S silent	5048	3818	6387	6438	1986	928
23403 S D614G	7353	4798	8136	9219	2451	1080
25563 orf3a Q57H	2035	2544	7881	2755	915	700
26333 E T30I	2603	2930	1117	3871	543	530
28144 orf8 S84L	4200	1808	219	6717	957	889
	VSP0088-1	VSP0065-1	VSP0065-2	VSP0089-1	VSP0089-2	VSP0318-1

Base change

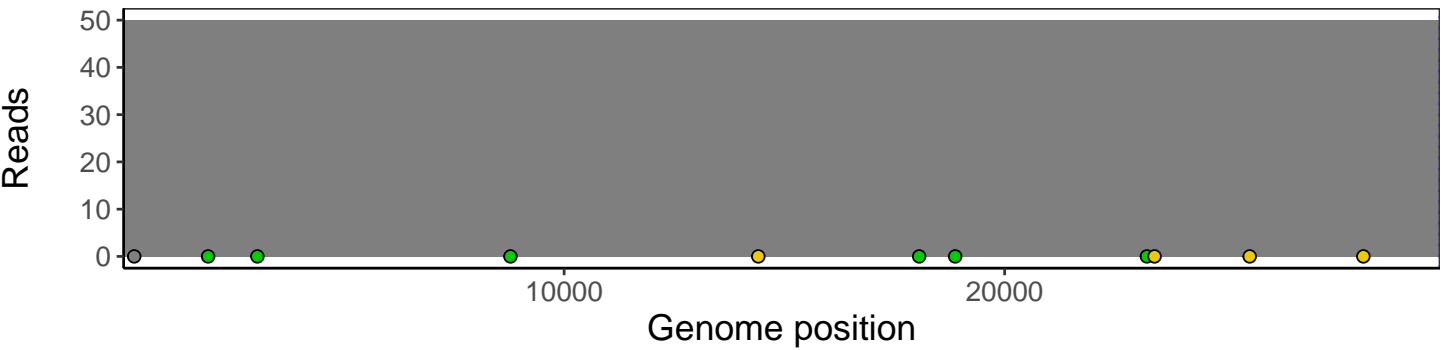- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

3

Analyses of individual experiments and composite results

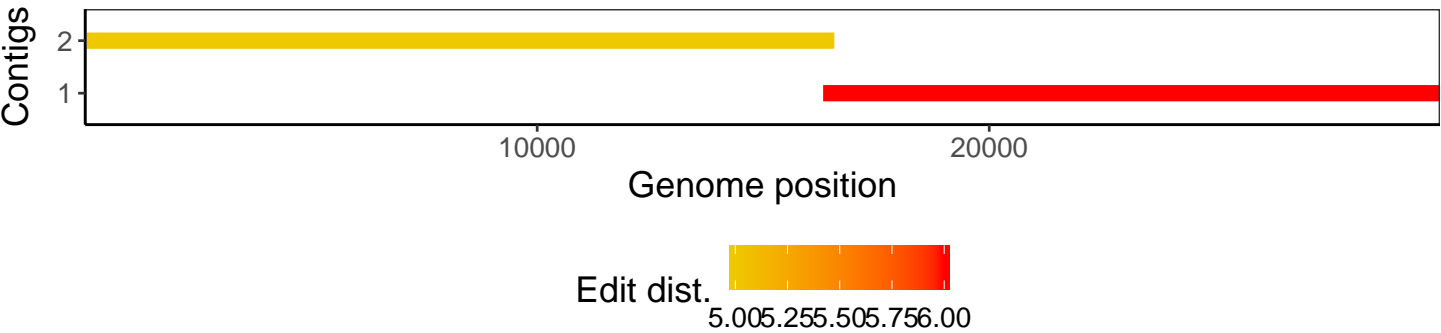## VSP0065 | 2020-05-04 | NP-OP | 251-q | composite result

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.
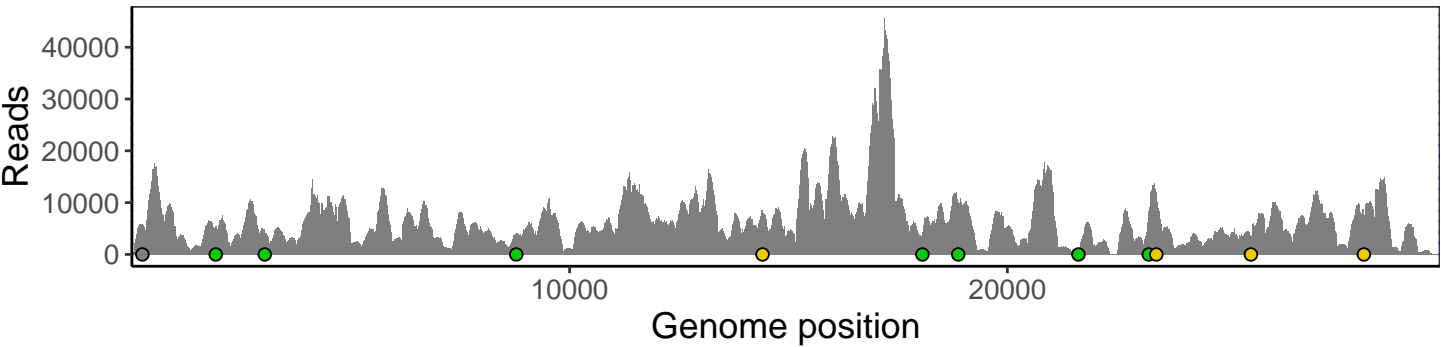

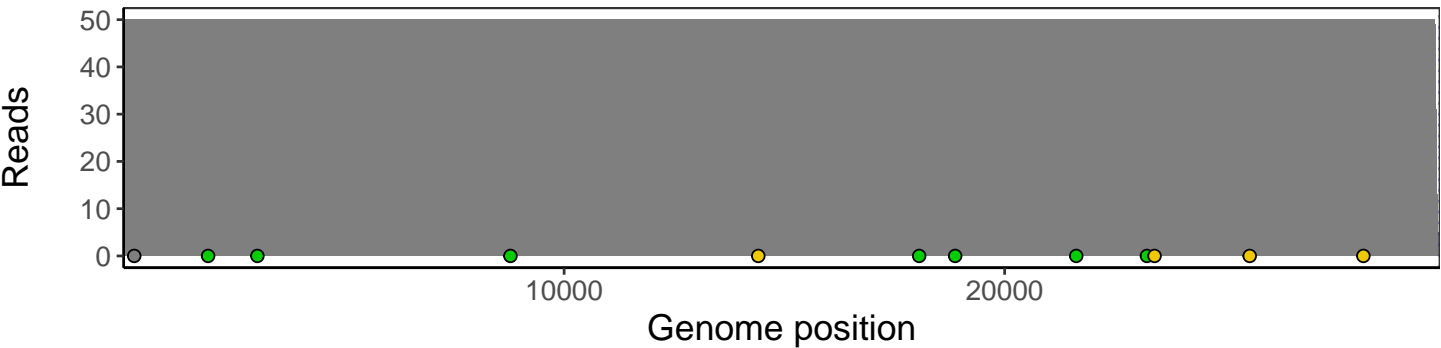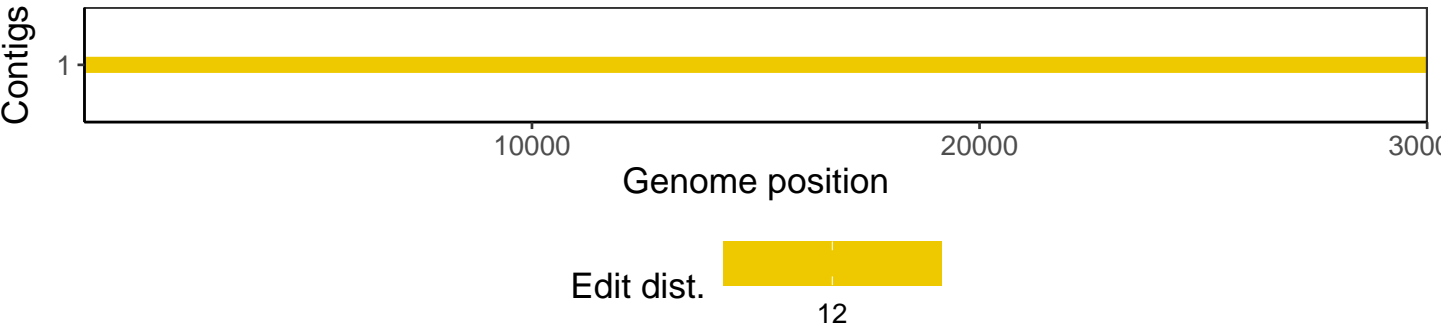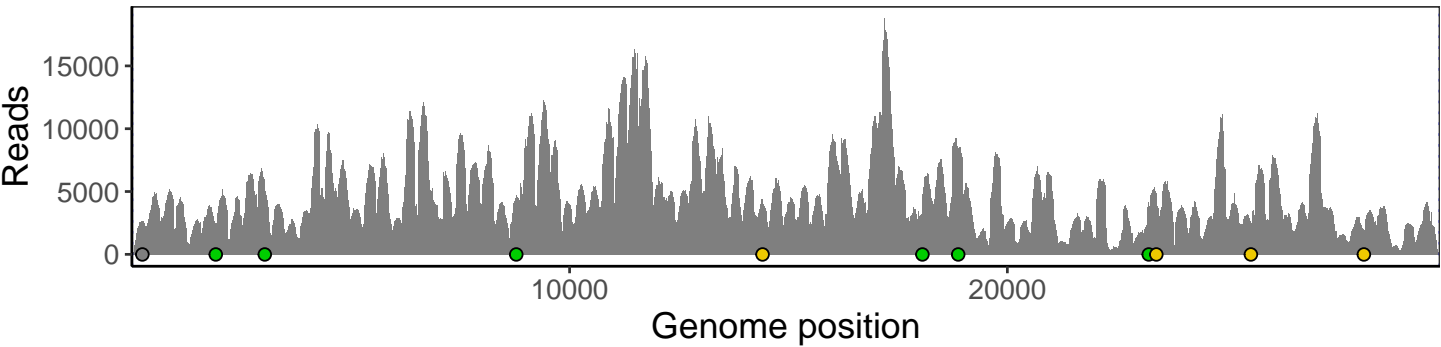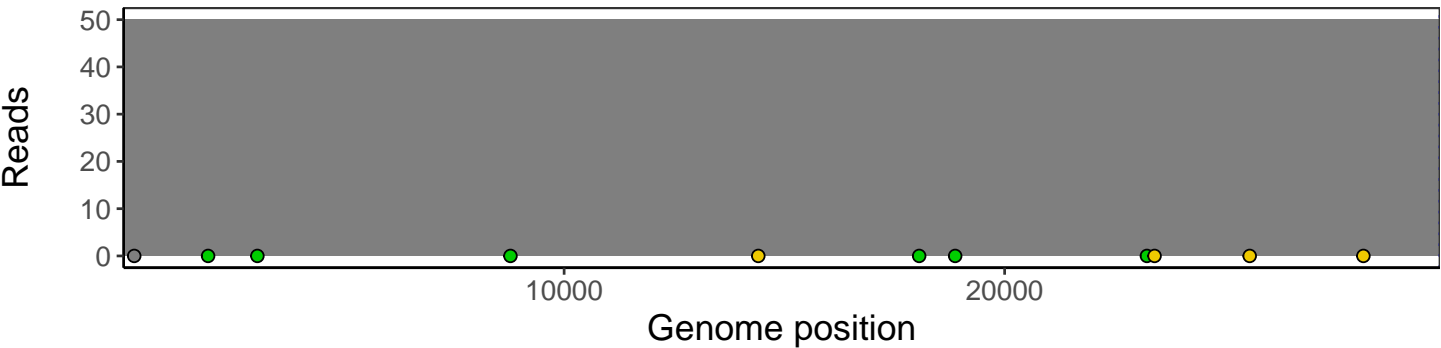Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

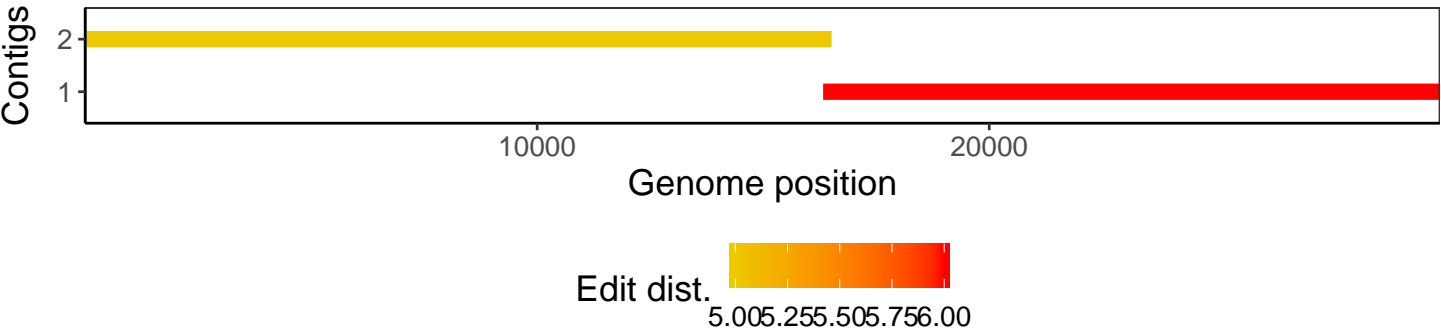**VSP0089 | 2020-05-06 | NP-OP | 251no-q | composite result**

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.
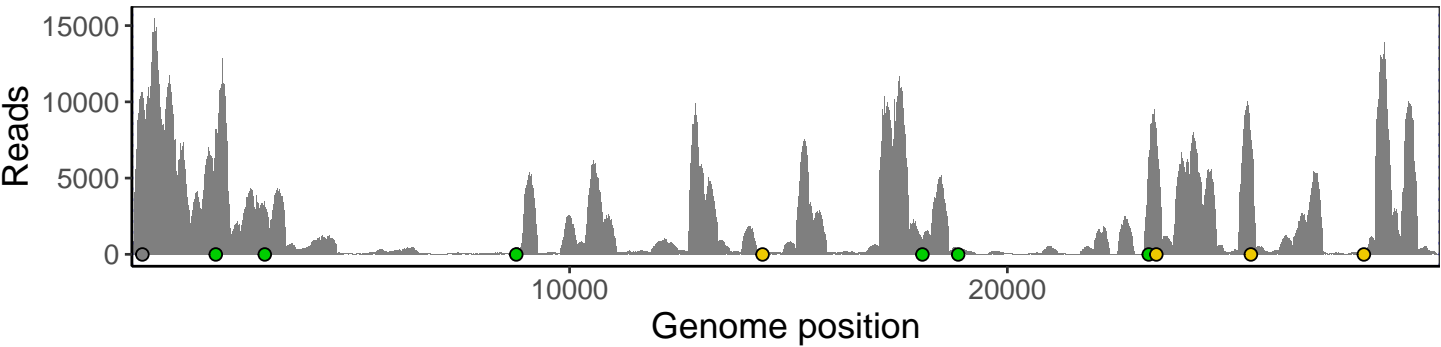

Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

VSP0065-1 | 2020-05-04 | NP-OP | 251-q | 7550000 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.

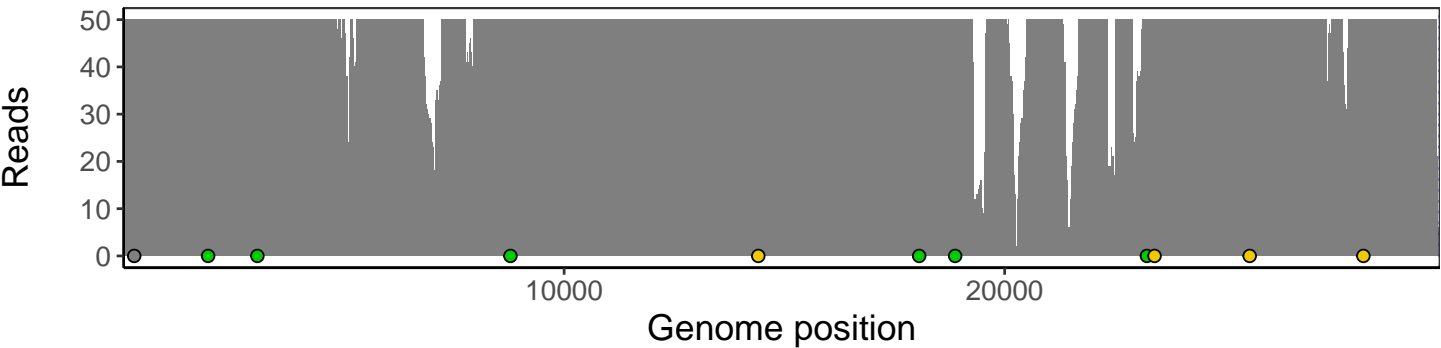Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

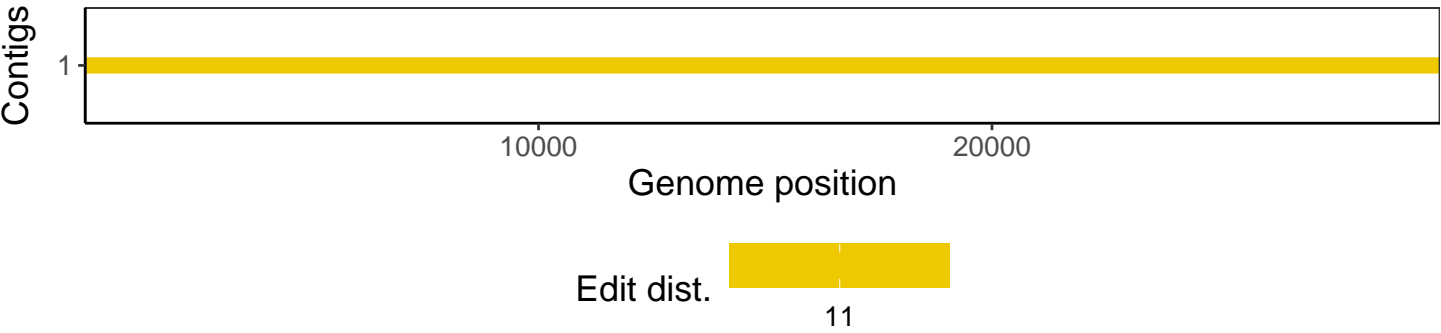**VSP0065-2 | 2020-05-04 | NP-OP | 251-q | 7550000 genomes | single experiment**

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.
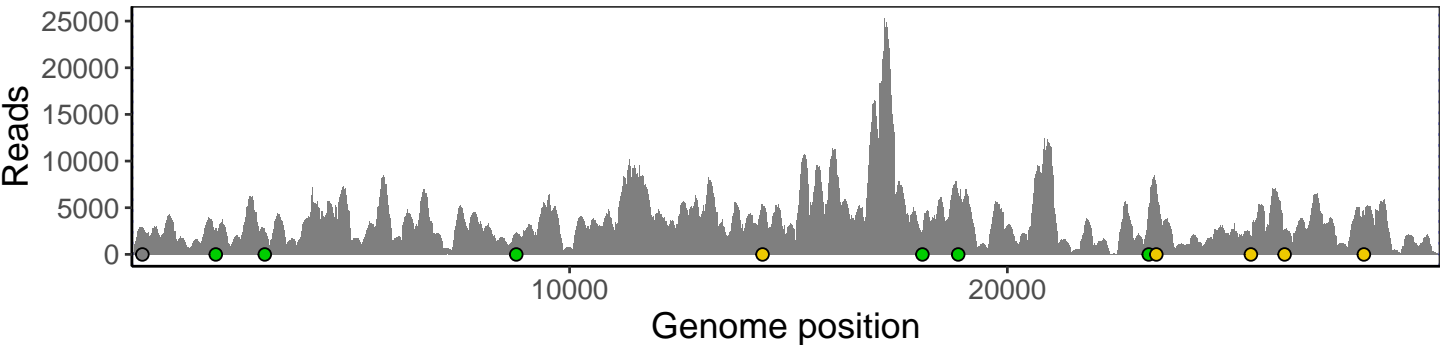

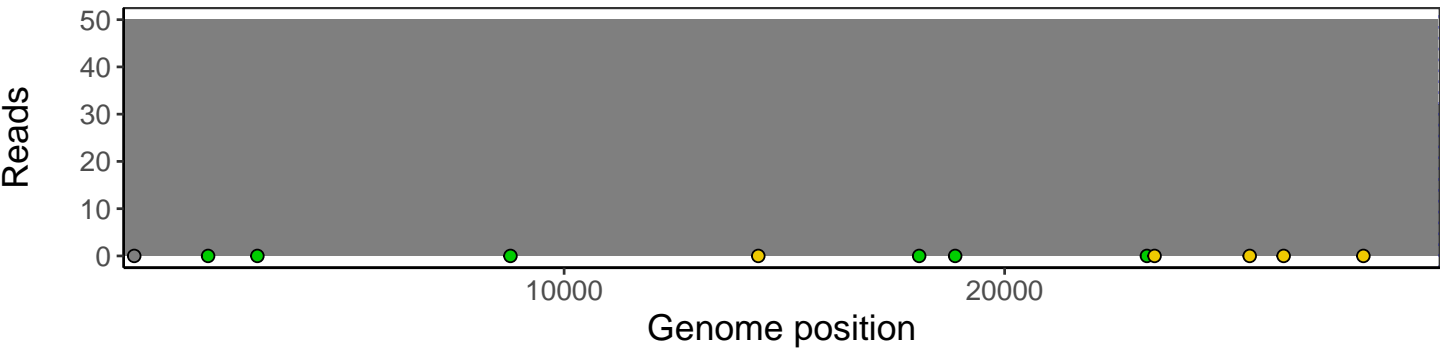Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

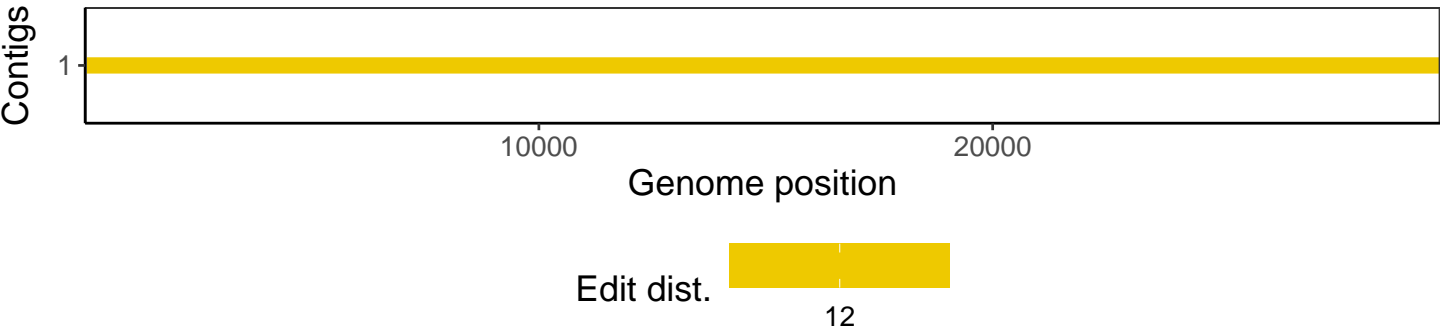**VSP0088-1 | 2020-05-06 | ETA | 251e-q | 255500 genomes | single experiment**

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.
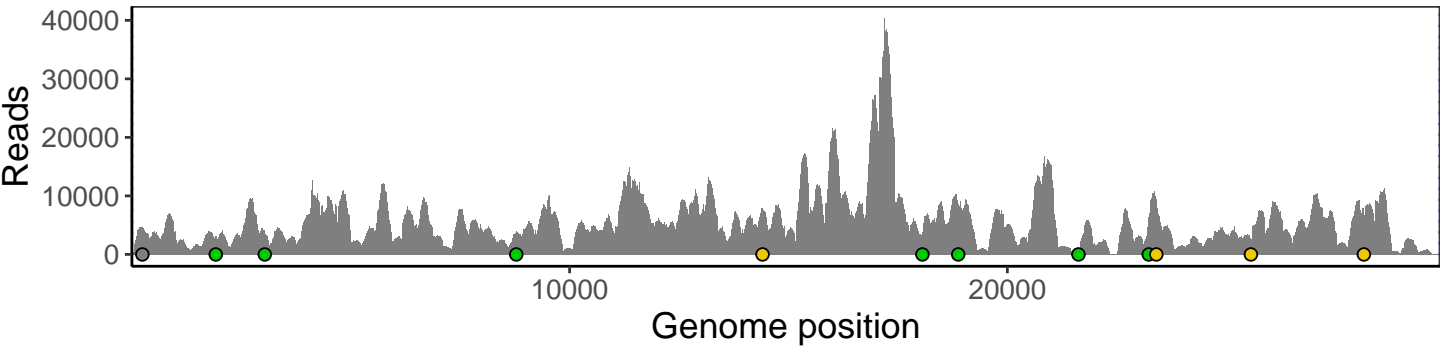

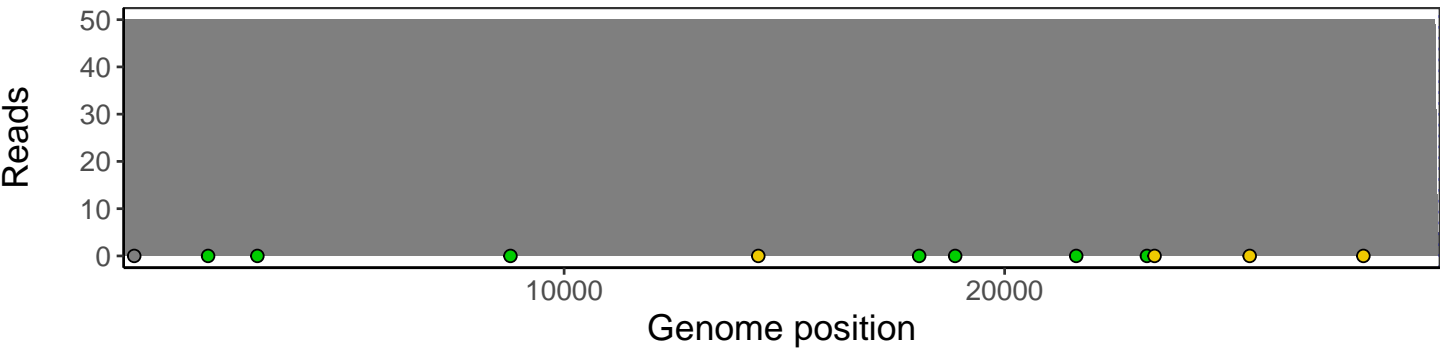Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

VSP0089-1 | 2020-05-06 | NP-OP | 251no-q | 570000 genomes | single experiment
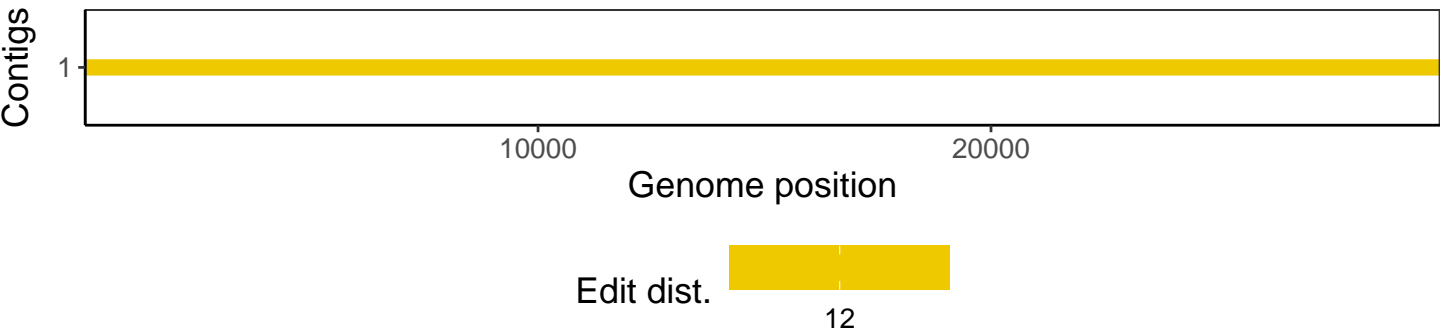
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.
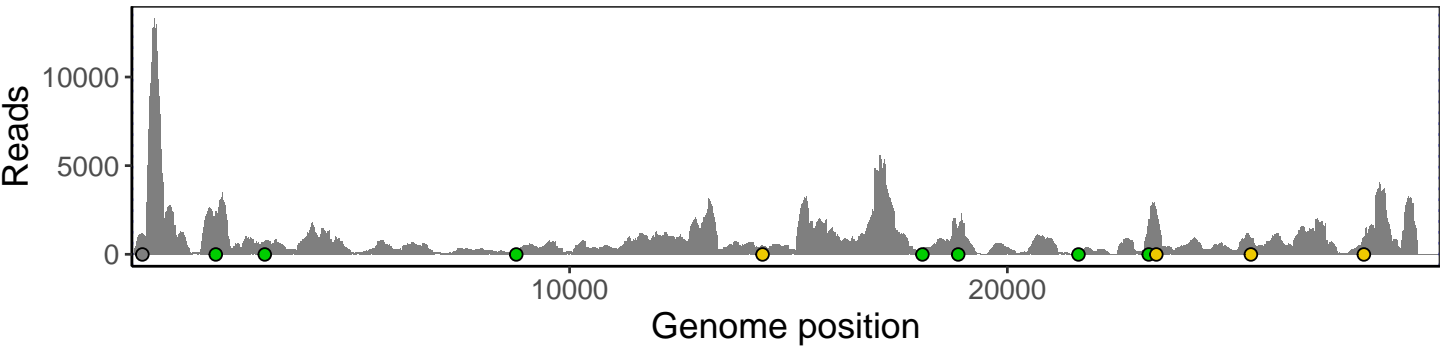

Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

VSP0089-2 | 2020-05-06 | NP-OP | 251no-q | 570000 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.

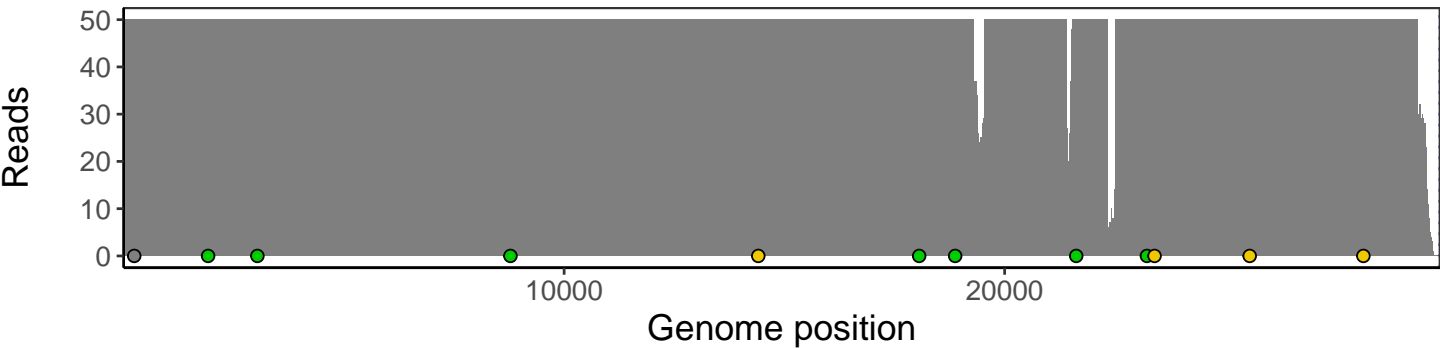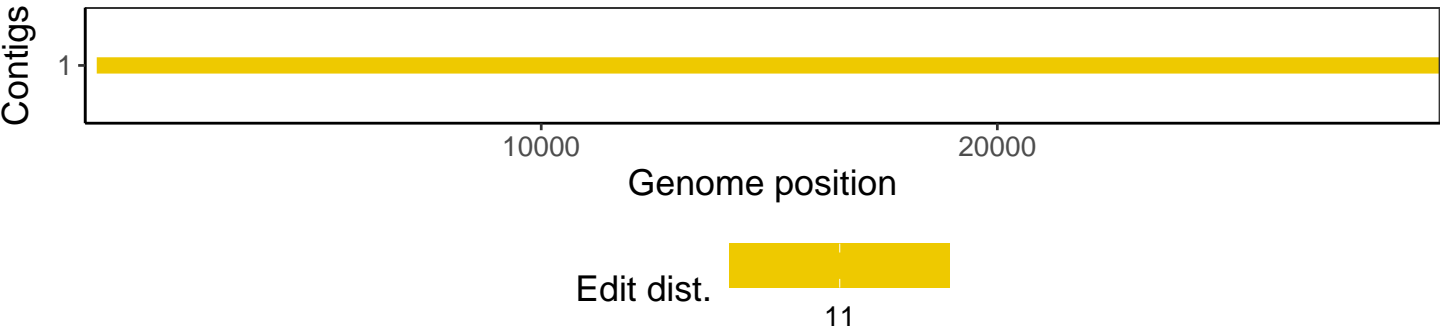Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

VSP0318-1 | 2020-05-06 | NP-OP | 251no-q | 890000 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.
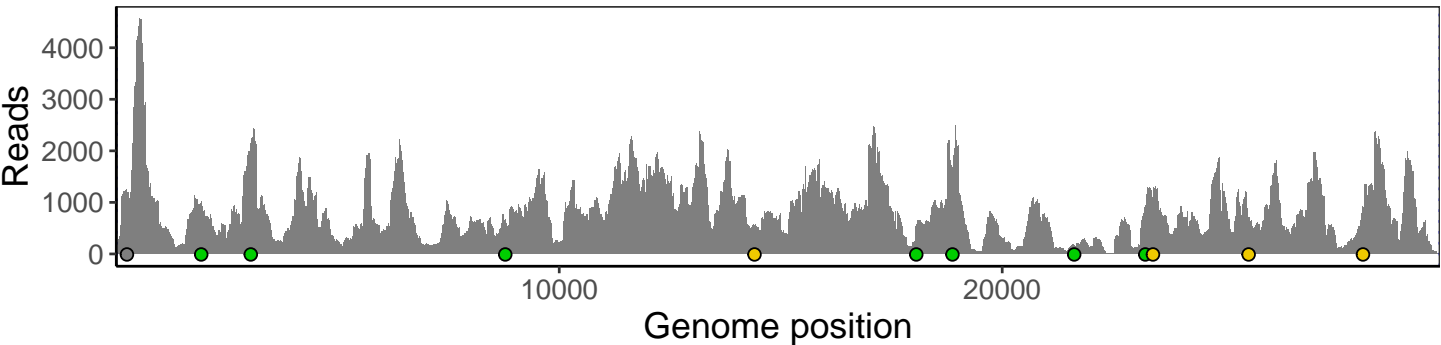

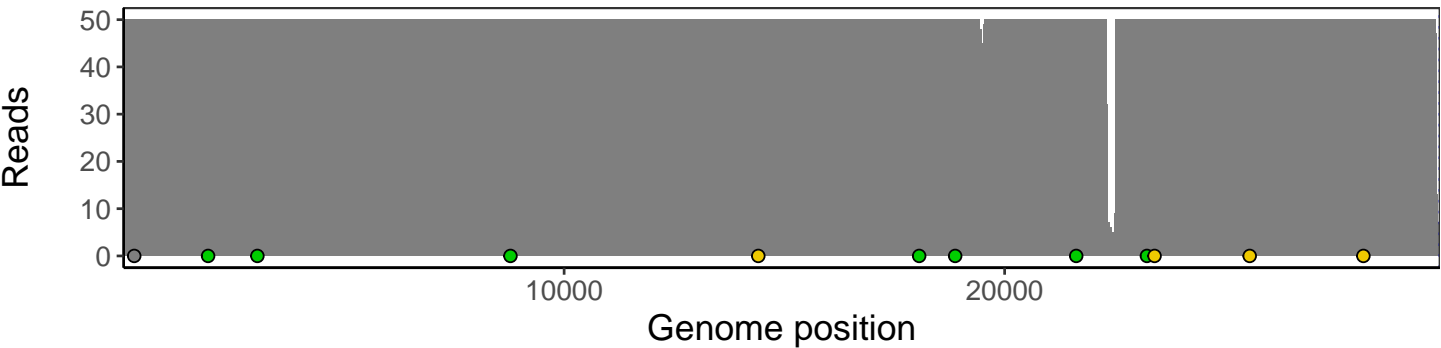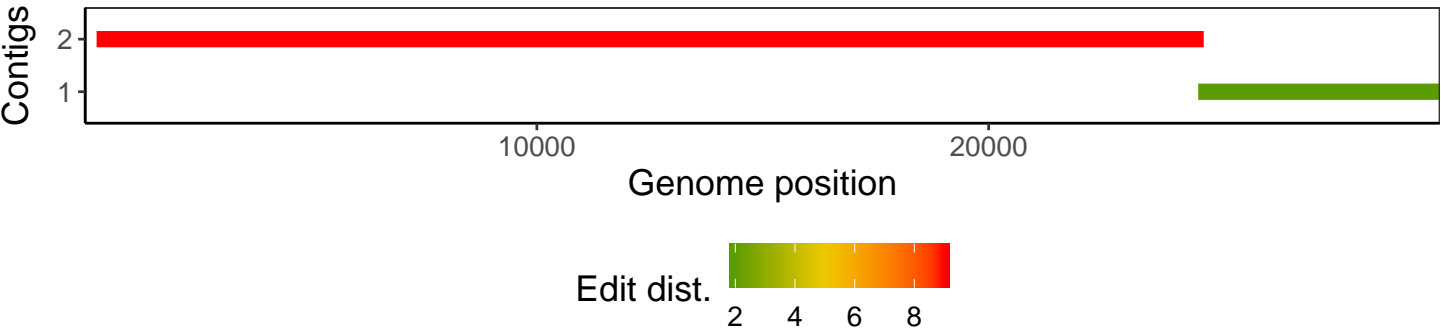Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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	2.1.7
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