COVID-19 subject 222-TCE

2021-01-28

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)
VSP0181	composite	NA	OP	2020-04-17	29.34	B.1.3	99.9%	99.9%
VSP0182	composite	NA	OP	2020-04-20	29.57	B.1.3	99.9%	99.9%
VSP0180-1m	single experiment	NA	NP	2020-04-17	29.87	B.1.3	99.9%	99.9%
VSP0181-1m	single experiment	NA	OP	2020-04-17	30.04	B.1.3	99.9%	99.9%
VSP0181-2	single experiment	1.510e+11	OP	2020-04-17	29.34	B.1.3	99.9%	99.8%
VSP0182-1m	single experiment	NA	OP	2020-04-20	29.87	B.1.3	99.9%	99.9%
VSP0182-2	single experiment	1.325e+11	OP	2020-04-20	20.33	B.1.3	99.9%	99.8%

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.
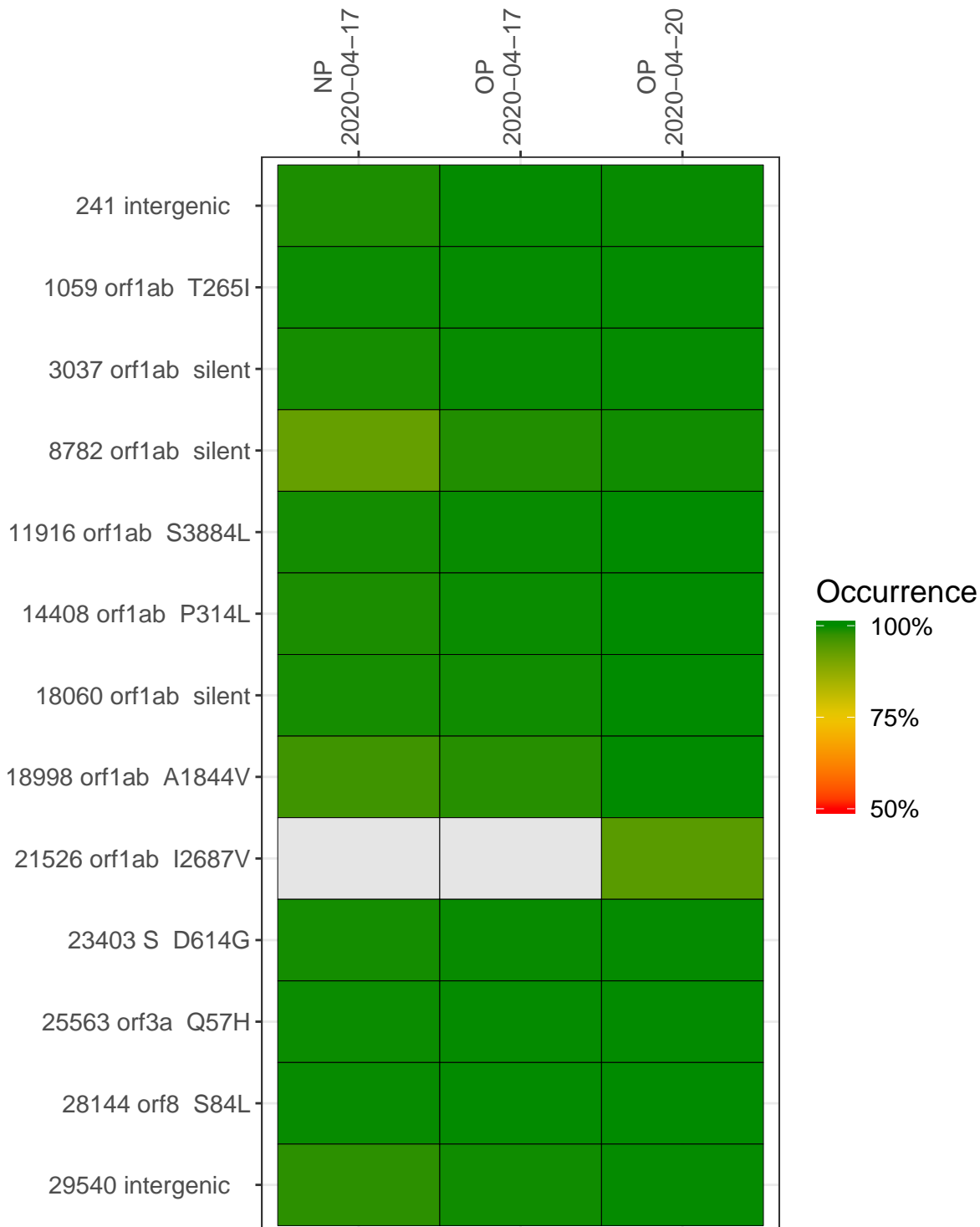
	NP 2020-04-17	OP 2020-04-17		OP 2020-04-20	
	VSP0180-1m	VSP0181-1m	VSP0181-2	VSP0182-1m	VSP0182-2
241 intergenic	1538	1426	1131	1527	1421
1059 orf1ab T265I	6929	10468	1073	5845	1280
3037 orf1ab silent	2591	7061	1413	3075	1330
8782 orf1ab silent	218	581	960	585	429
11916 orf1ab S3884L	1213	4748	480	1911	221
14408 orf1ab P314L	1871	6510	926	2536	235
18060 orf1ab silent	1572	6318	774	2324	492
18998 orf1ab A1844V	176	509	725	660	266
21526 orf1ab I2687V	30	72	61	73	27
23403 S D614G	3058	3898	2005	3688	2063
25563 orf3a Q57H	5434	9712	1538	5364	2005
28144 orf8 S84L	682	1896	570	1289	382
29540 intergenic	320	1600	603	620	506

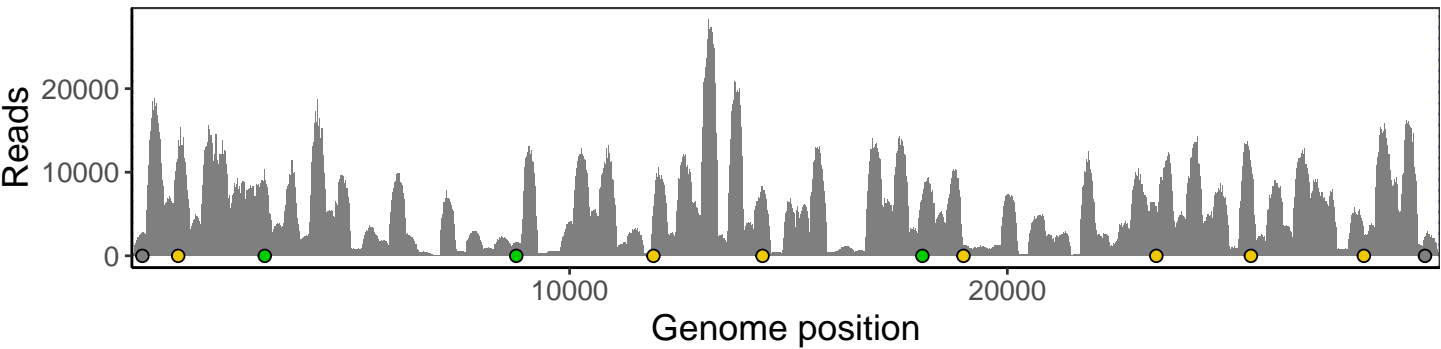Base change

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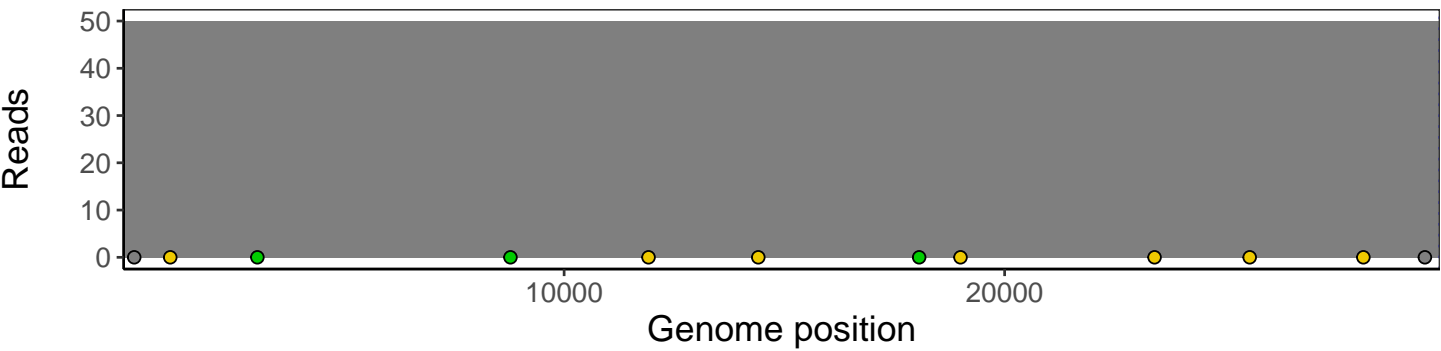- A
- T
- C
- G
- N
- Ins/Del
- No data

3

Analyses of individual experiments and composite results

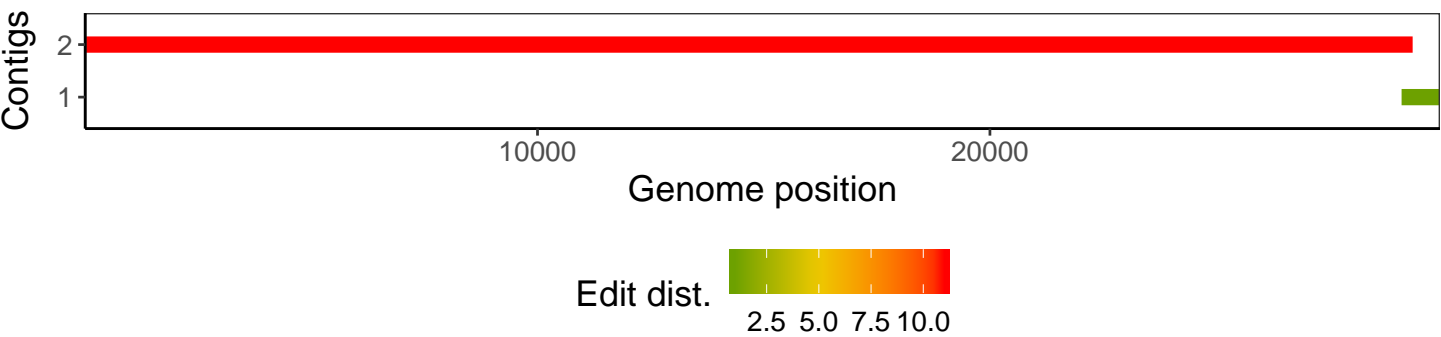**VSP0181 | 2020-04-17 | OP | 2 | composite result**

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 by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.
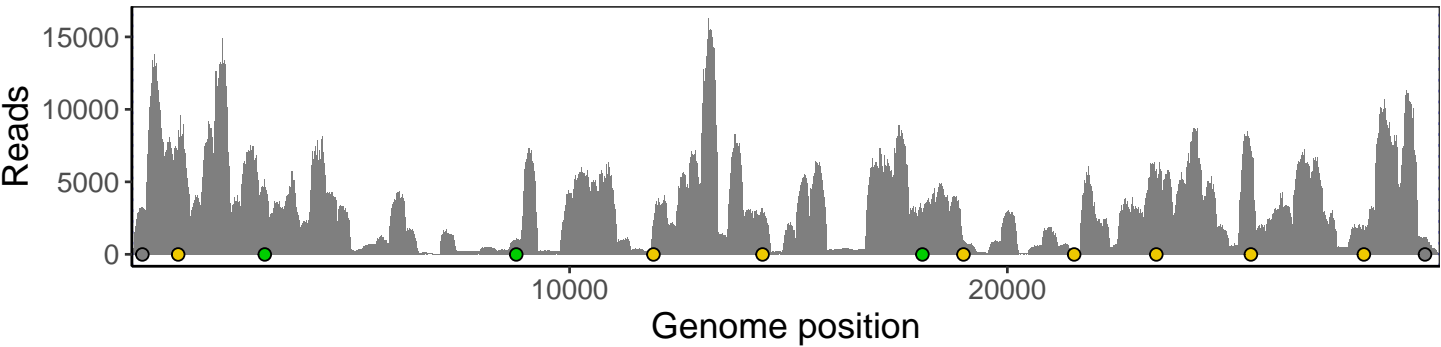

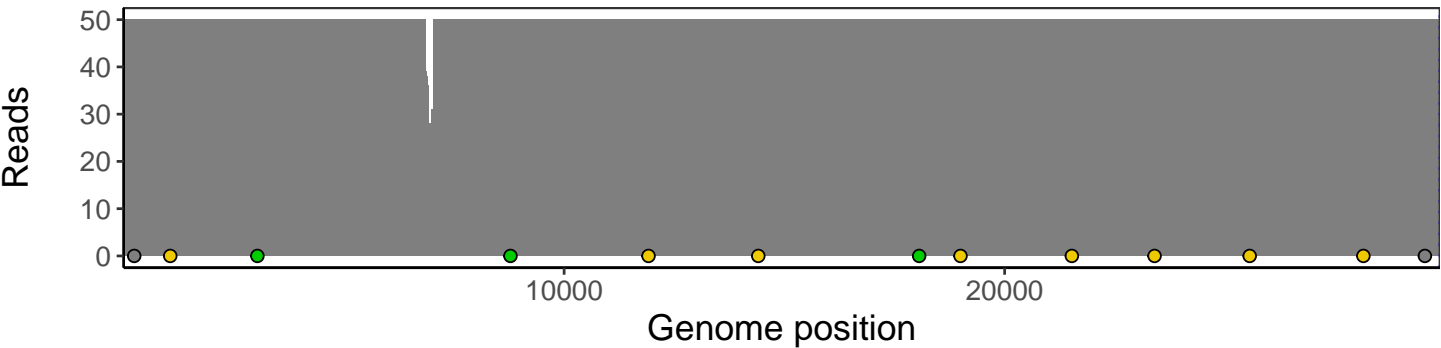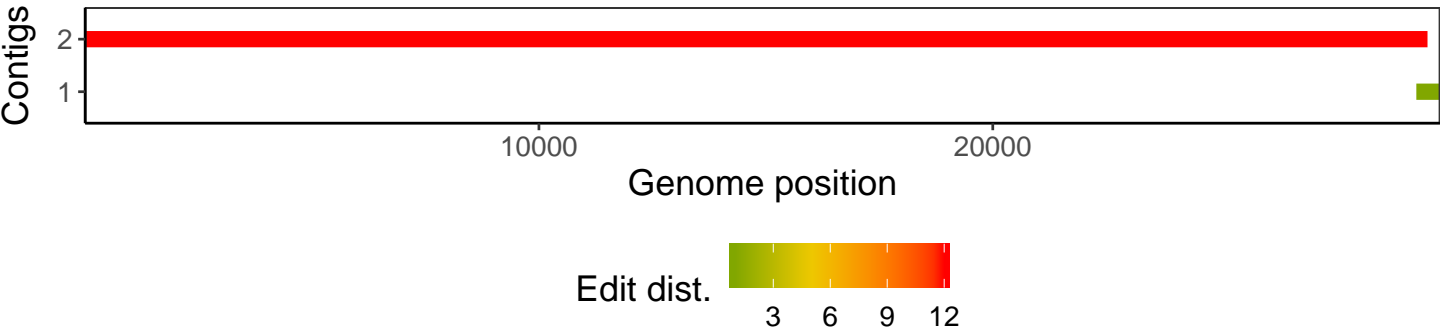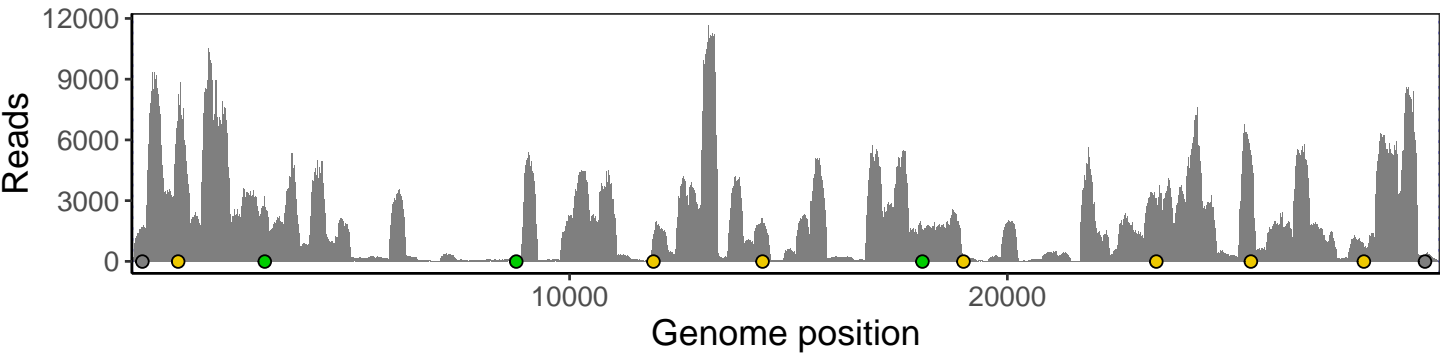Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

VSP0182 | 2020-04-20 | OP | 3 | 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.

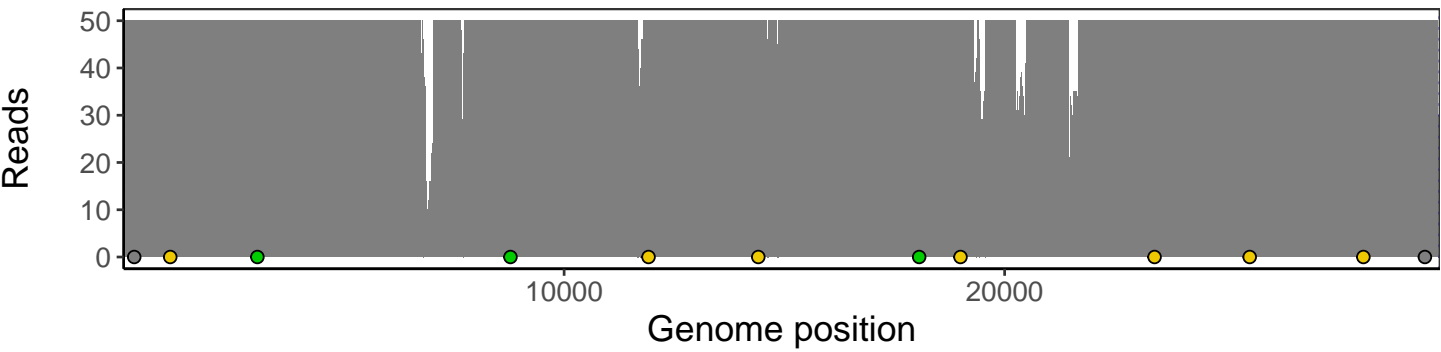Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

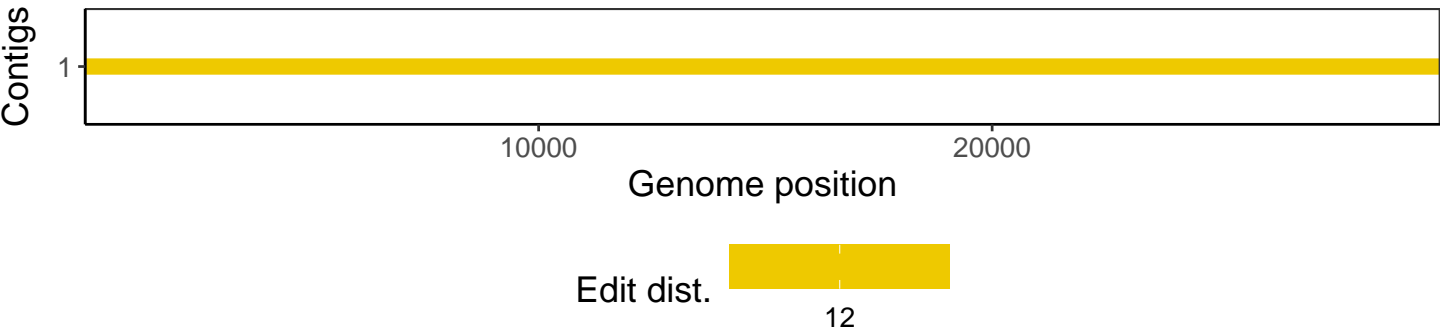**VSP0180-1m | 2020-04-17 | NP | 1 | genomes | single experiment**

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.
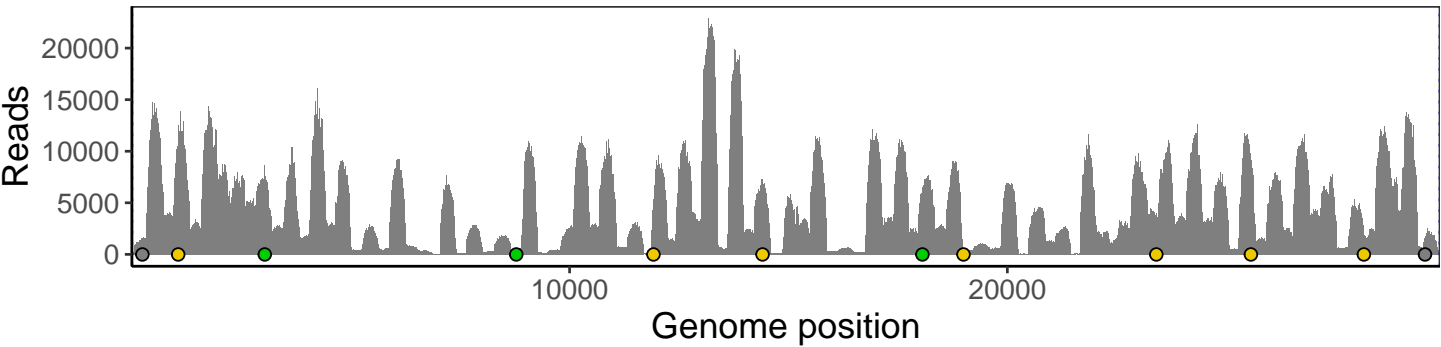

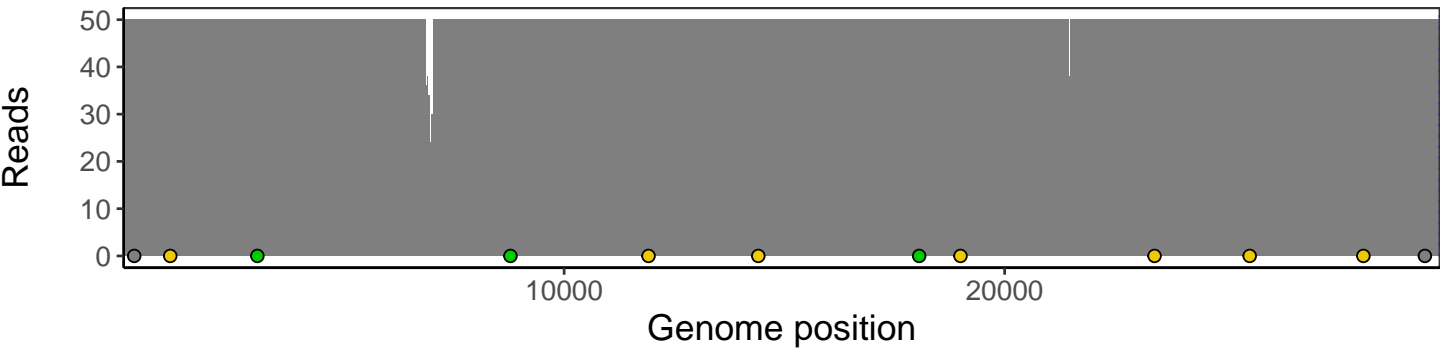Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

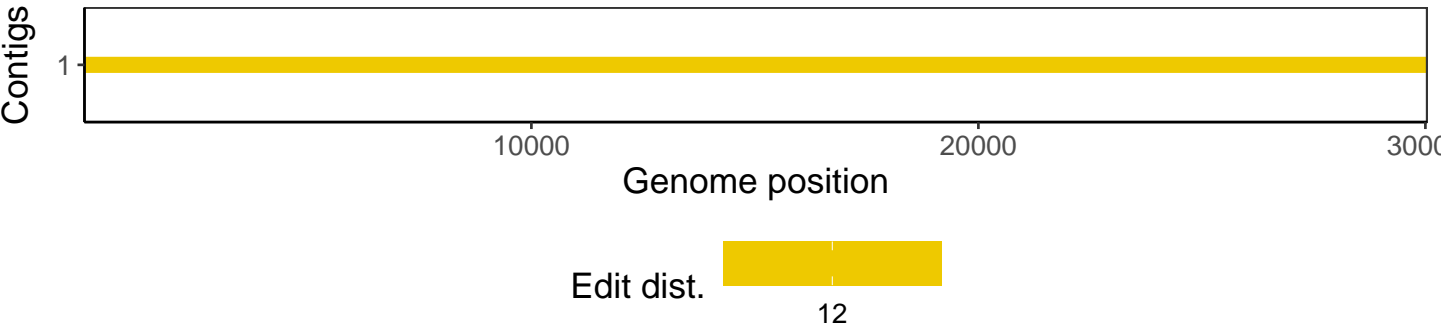**VSP0181-1m | 2020-04-17 | OP | 2 | genomes | single experiment**

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.
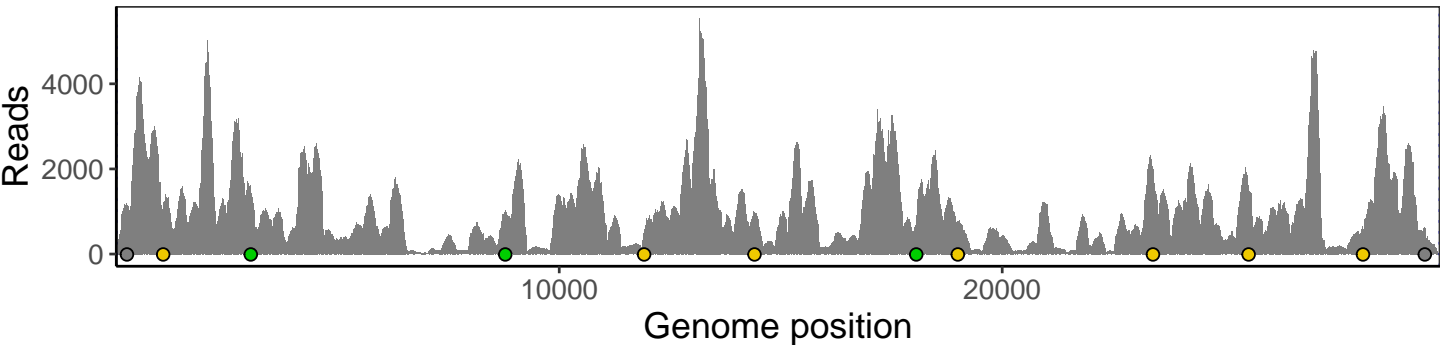

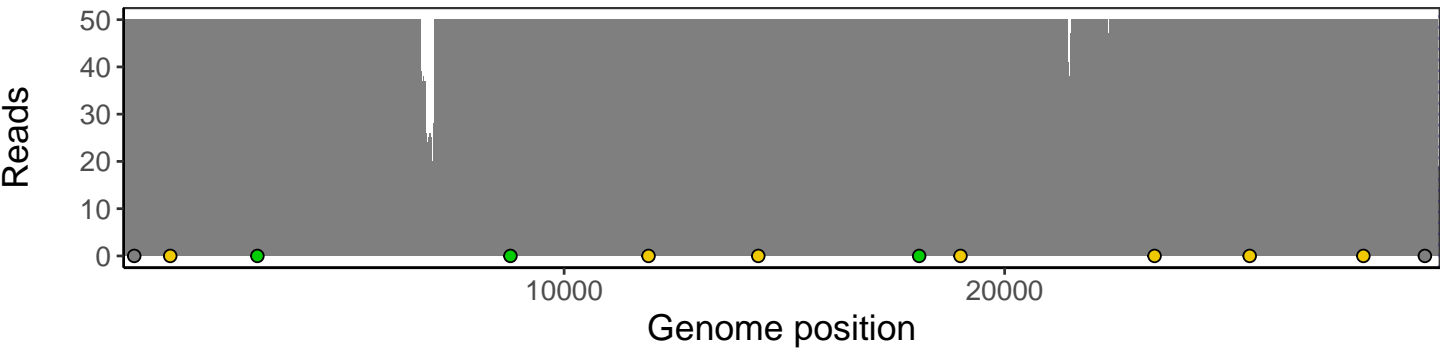Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

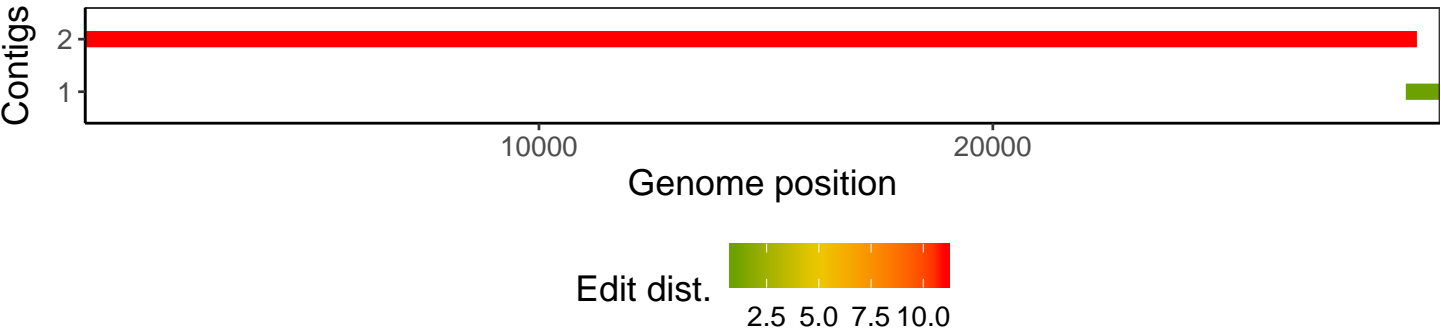**VSP0181-2 | 2020-04-17 | OP | 2 | 1.51e+11 genomes | single experiment**

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.
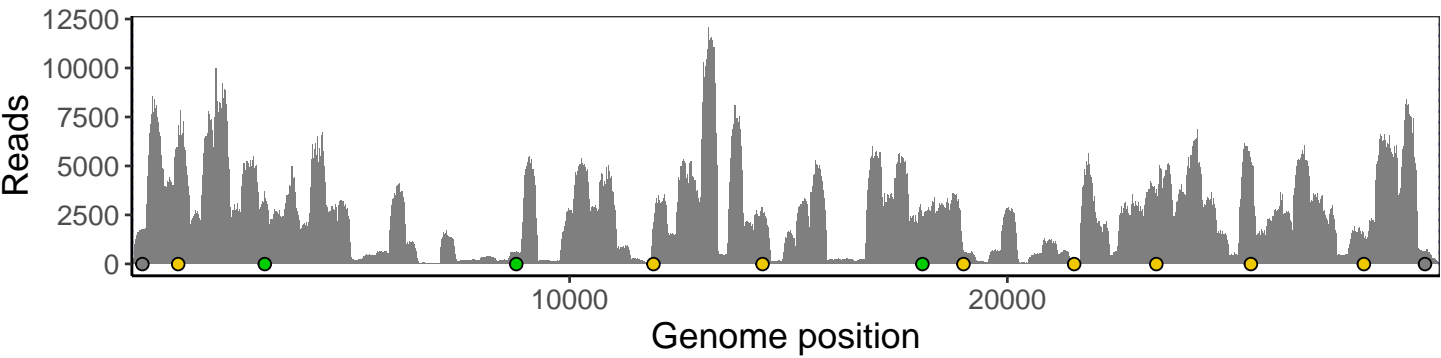

Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

8

VSP0182-1m | 2020-04-20 | OP | 3 | 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.

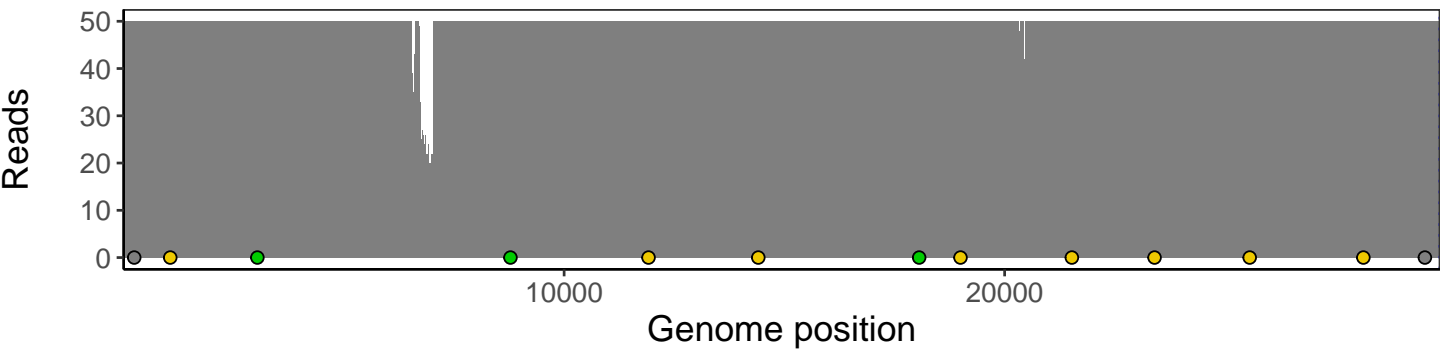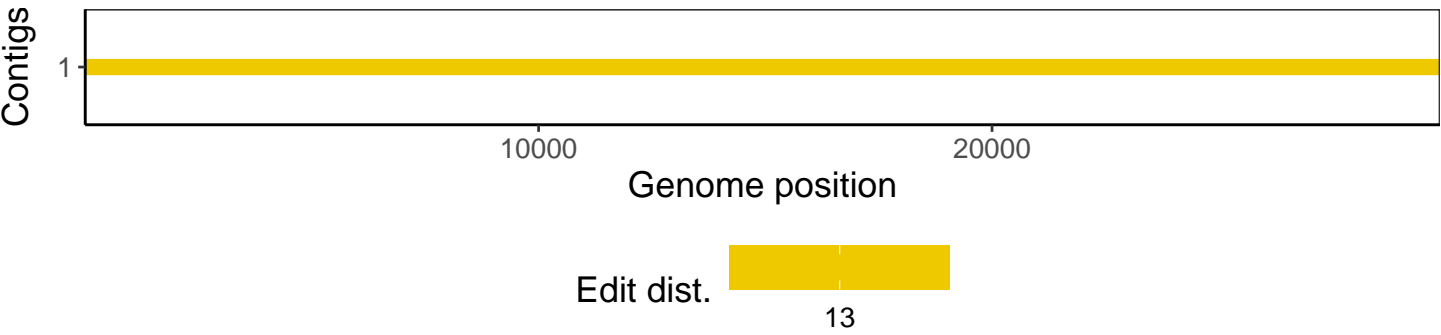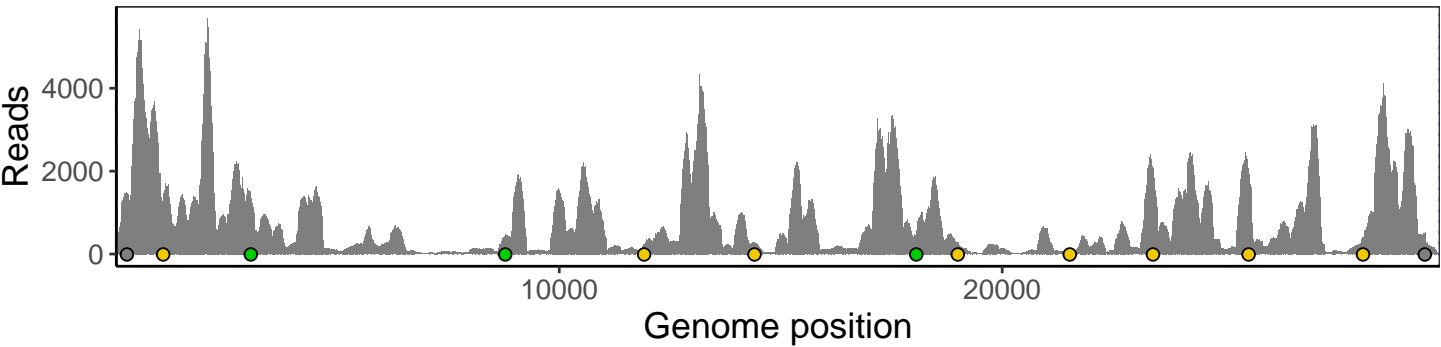Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

VSP0182-2 | 2020-04-20 | OP | 3 | 1.325e+11 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.
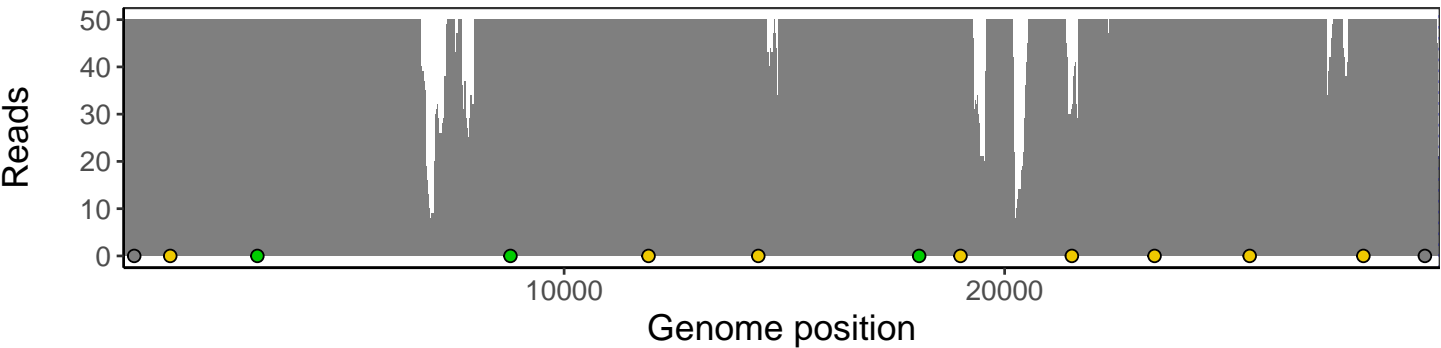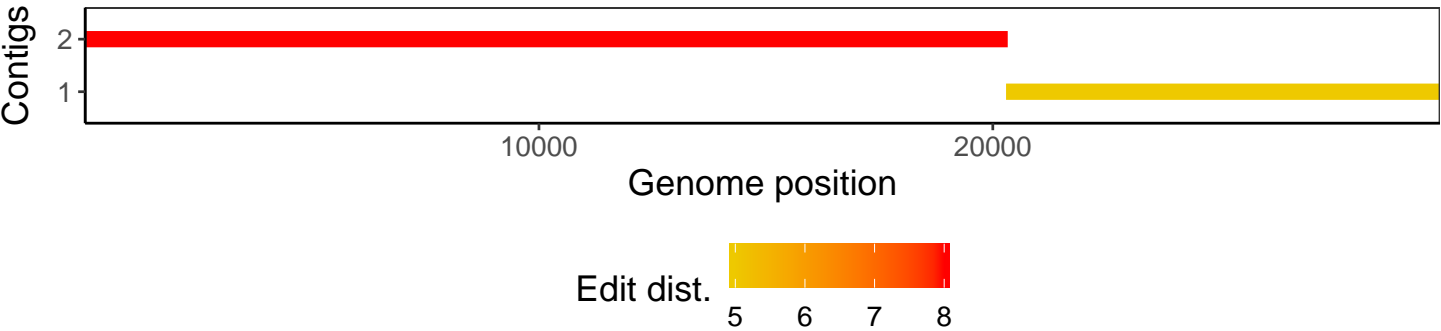

The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

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