COVID-19 subject 266

2020-09-14

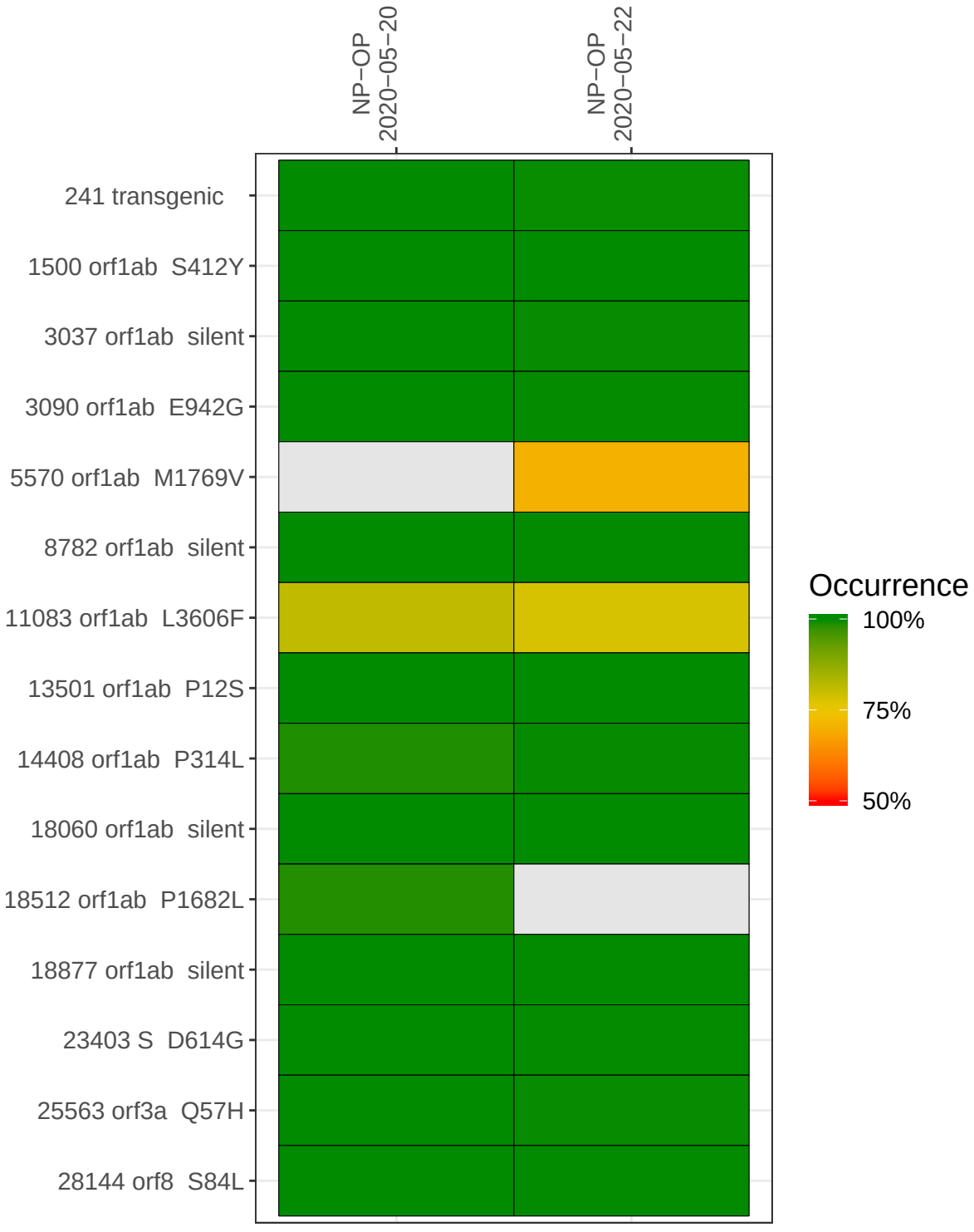
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. The code base for this analysis can be found (here).

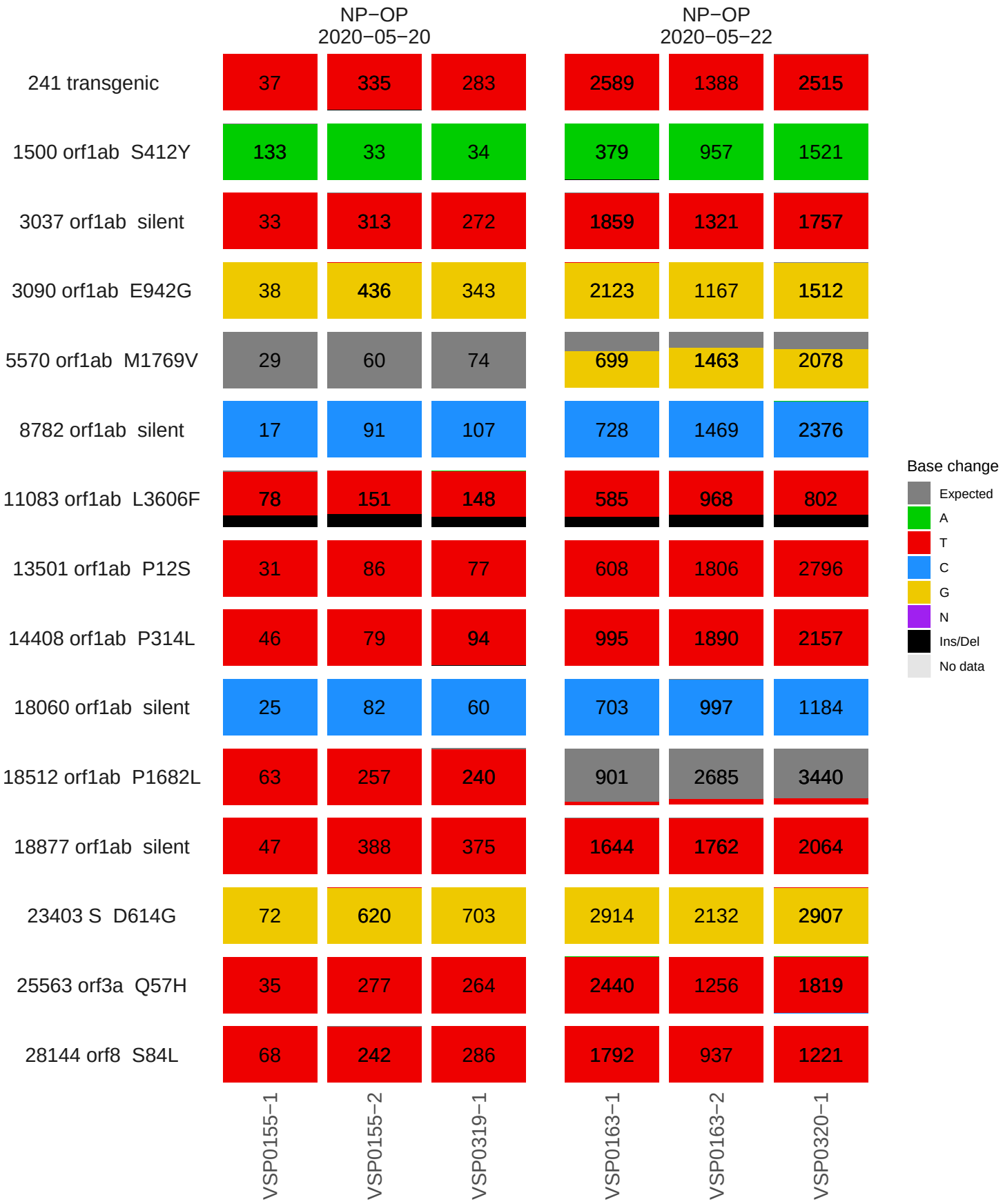
Table 1. Sample summary.

Experiment	Type	Input genomes	Sample type	Sample date	Largest contig (KD)	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0155	composite	NA	NP-OP	2020-05-20	22.33	99.8%	99.0%
VSP0163	composite	NA	NP-OP	2020-05-22	29.75	99.9%	99.8%
VSP0155-1	single experiment	17000	NP-OP	2020-05-20	21.42	99.4%	98.2%
VSP0155-2	single experiment	85000	NP-OP	2020-05-20	22.33	99.7%	98.6%
VSP0163-1	single experiment	1530000	NP-OP	2020-05-22	29.75	99.9%	99.8%
VSP0163-2	single experiment	7650000	NP-OP	2020-05-22	29.84	99.9%	99.8%
VSP0319-1	single experiment	143500	NP-OP	2020-05-20	22.25	99.3%	98.5%
VSP0320-1	single experiment	5750000	NP-OP	2020-05-22	29.86	99.8%	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 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.

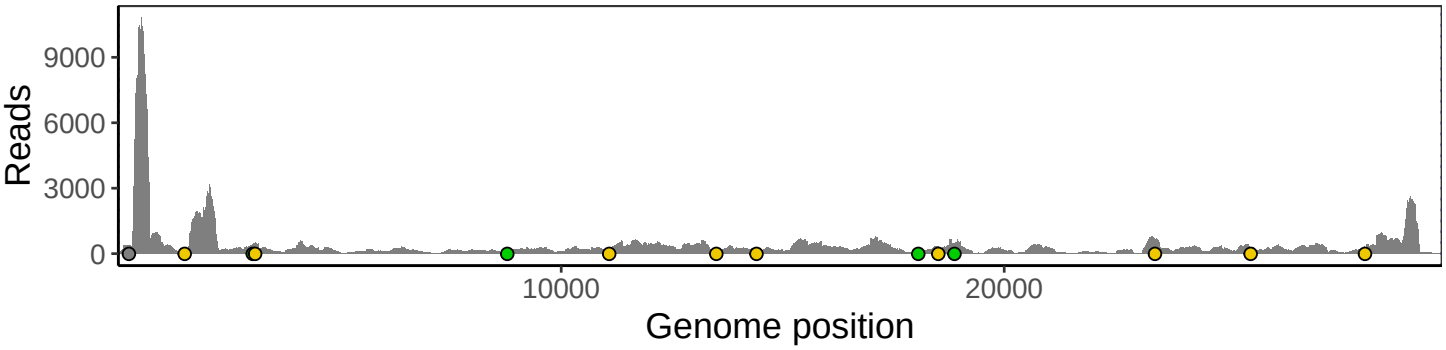



NP−OP
2020−05−20
NP−OP
2020−05−22
241 transgenic
1500 orf1ab S412Y
3037 orf1ab silent
3090 orf1ab E942G
5570 orf1ab M1769V
8782 orf1ab silent
11083 orf1ab L3606F
13501 orf1ab P12S
14408 orf1ab P314L
18060 orf1ab silent
18512 orf1ab P1682L
18877 orf1ab silent
23403 S D614G
25563 orf3a Q57H
28144 orf8 S84L
37 335 283 2589 1388 2515
133 33 34 379 957 1521
33 313 272 1859 1321 1757
38 436 343 2123 1167 1512
29 60 74 699 1463 2078
17 91 107 728 1469 2376
78 151 148 585 968 802
31 86 77 608 1806 2796
46 79 94 995 1890 2157
25 82 60 703 997 1184
63 257 240 901 2685 3440
47 388 375 1644 1762 2064
72 620 703 2914 2132 2907
35 277 264 2440 1256 1819
68 242 286 1792 937 1221
VSP0155−1
VSP0155−2
VSP0319−1
VSP0163−1
VSP0163−2
VSP0320−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

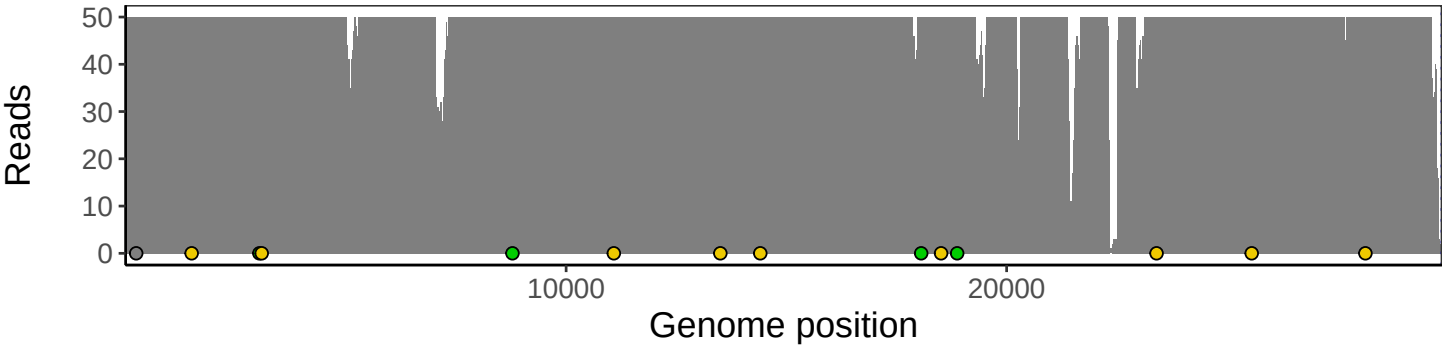
Analyses of individual experiments and composite results.

VSP0155 | 2020-05-20 | NP-OP | 266no-q | composite result

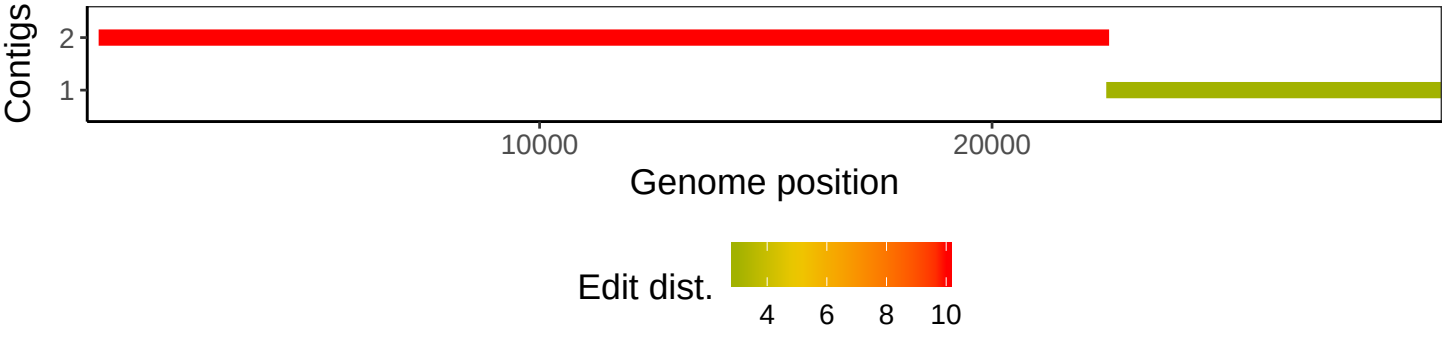
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

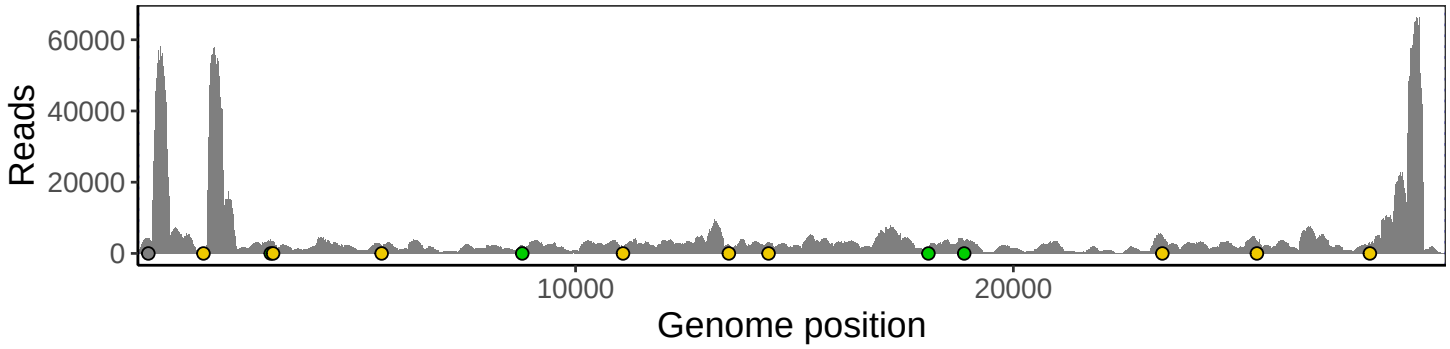


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

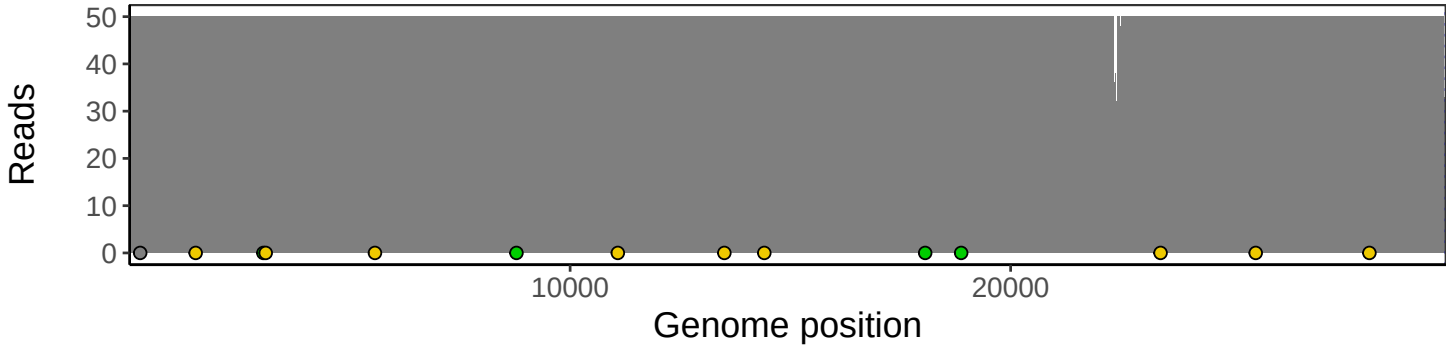


VSP0163 | 2020-05-22 | NP-OP | 266no-q | composite result

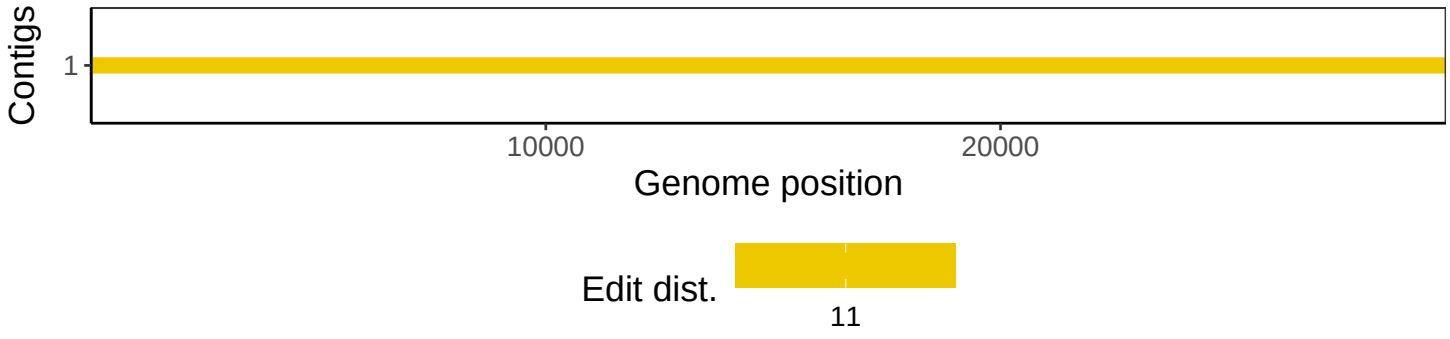
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

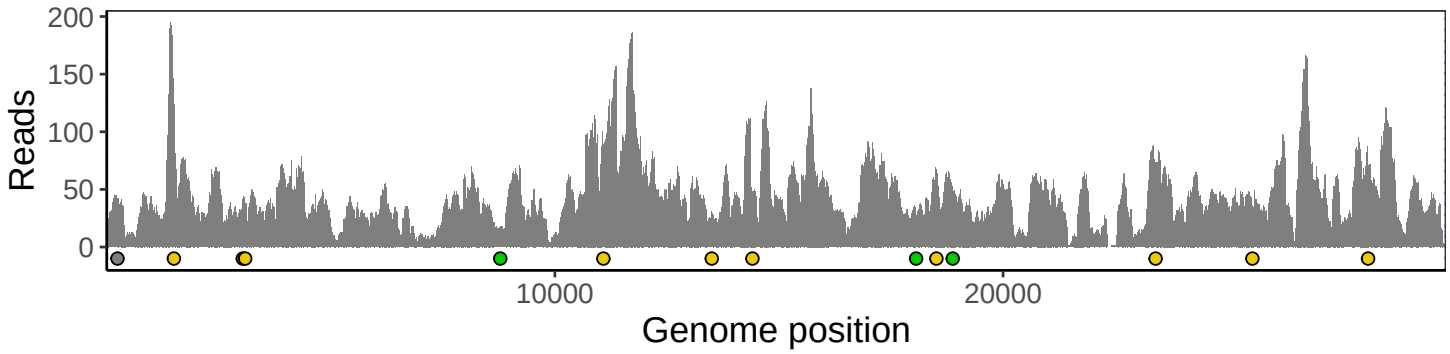


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

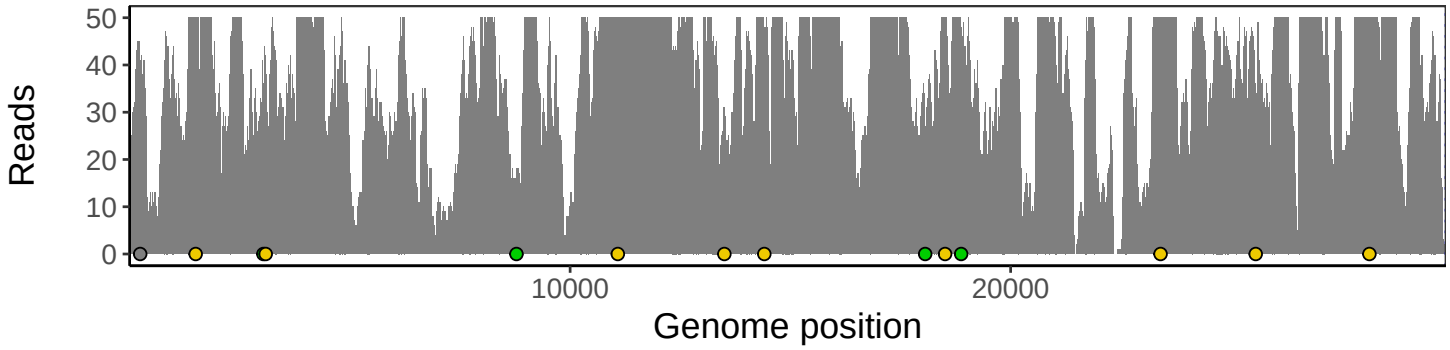


VSP0155-1 | 2020-05-20 | NP-OP | 266no-q | 17000 genomes | single experiment

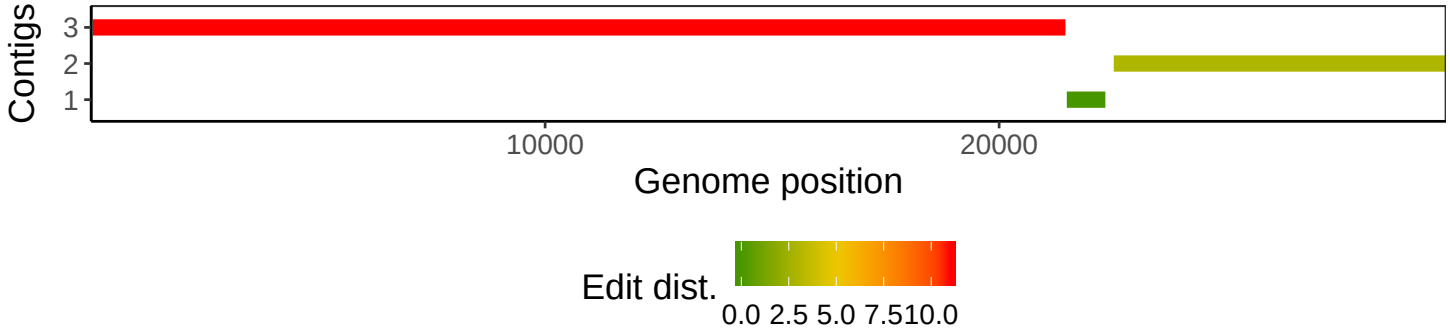
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

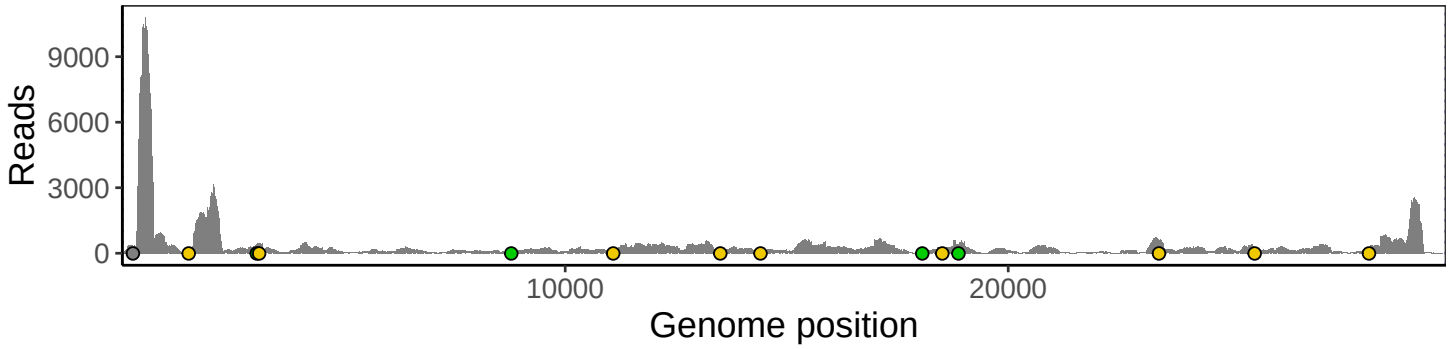


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

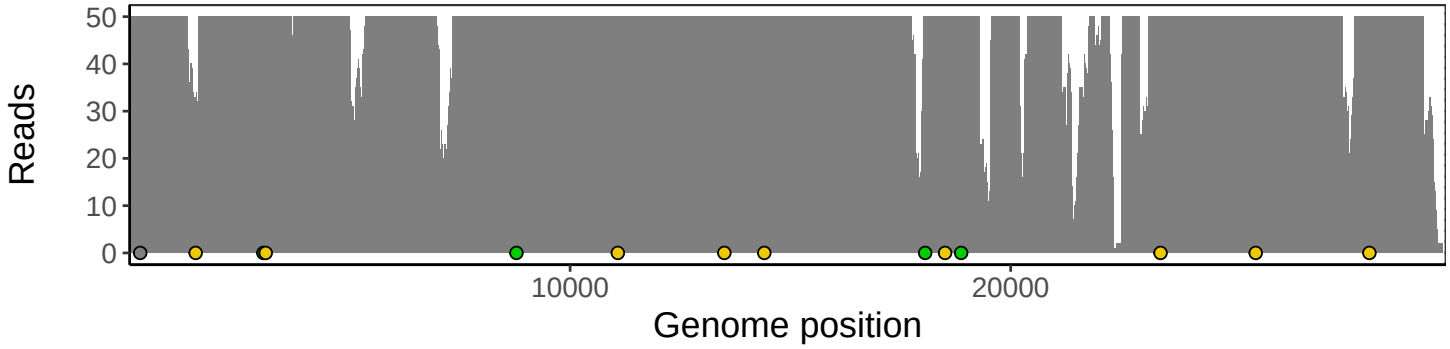


VSP0155-2 | 2020-05-20 | NP-OP | 266no-q | 85000 genomes | single experiment

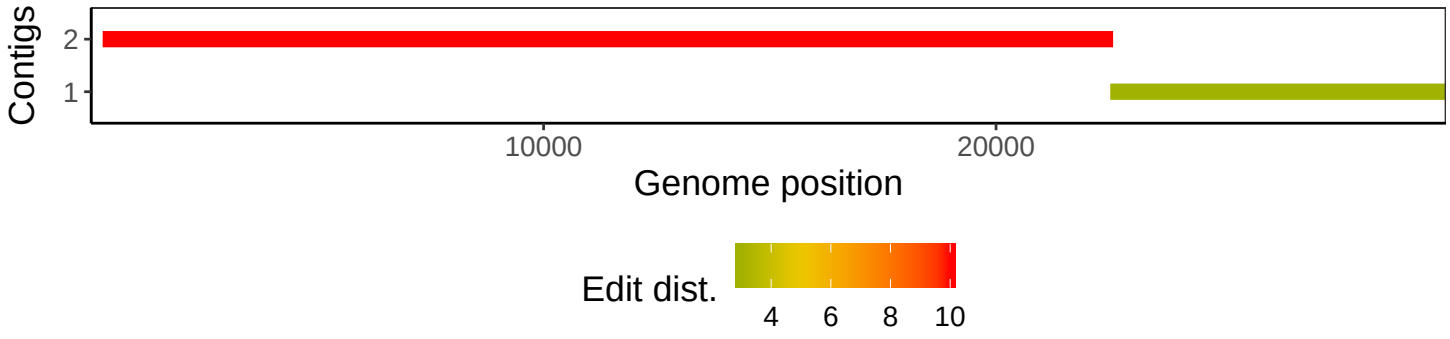
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

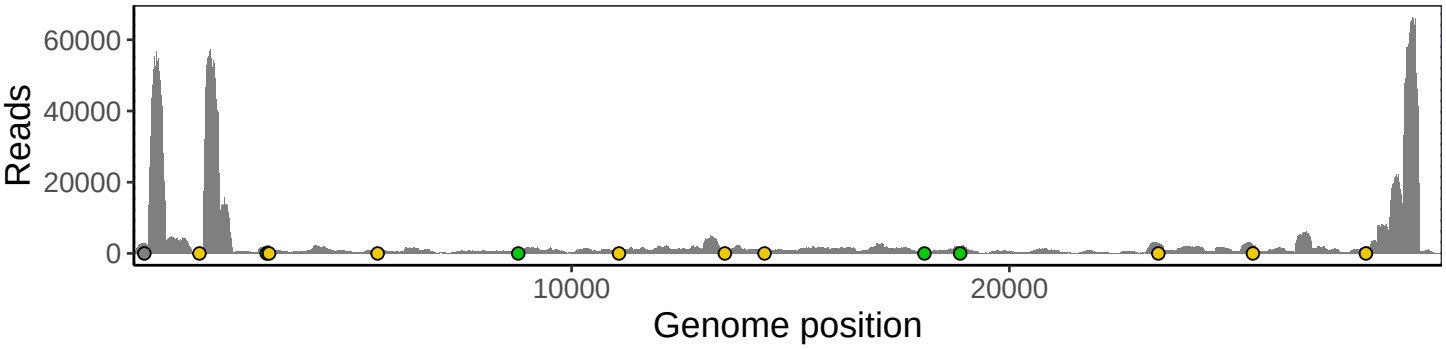


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

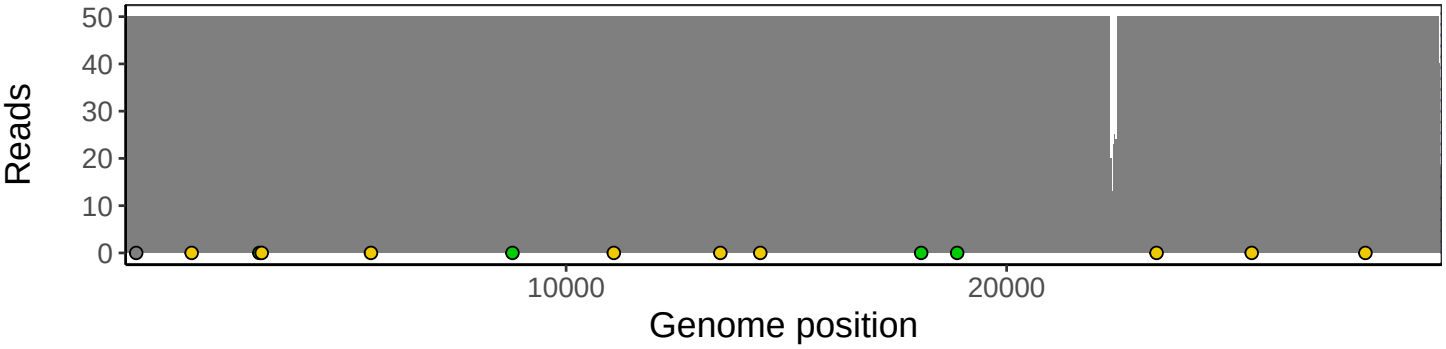


VSP0163-1 | 2020-05-22 | NP-OP | 266no-q | 1530000 genomes | single experiment

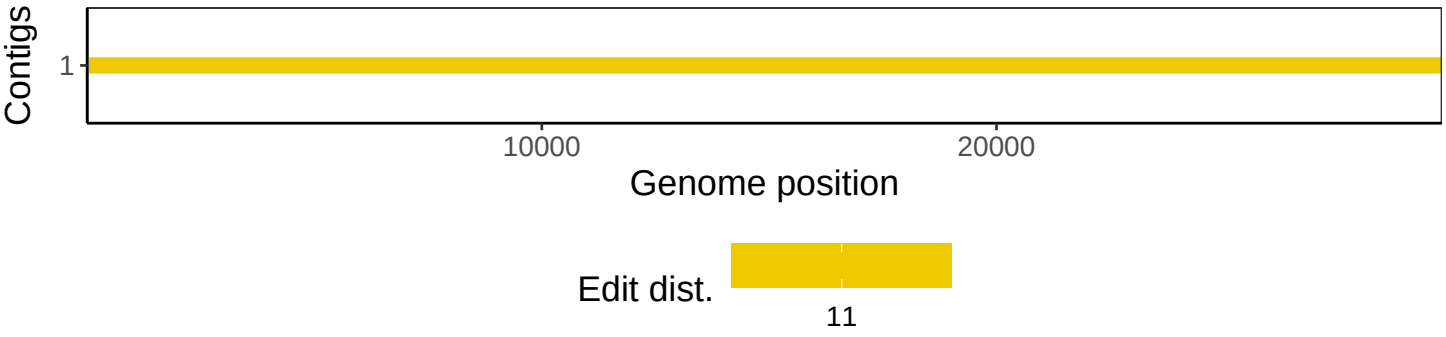
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

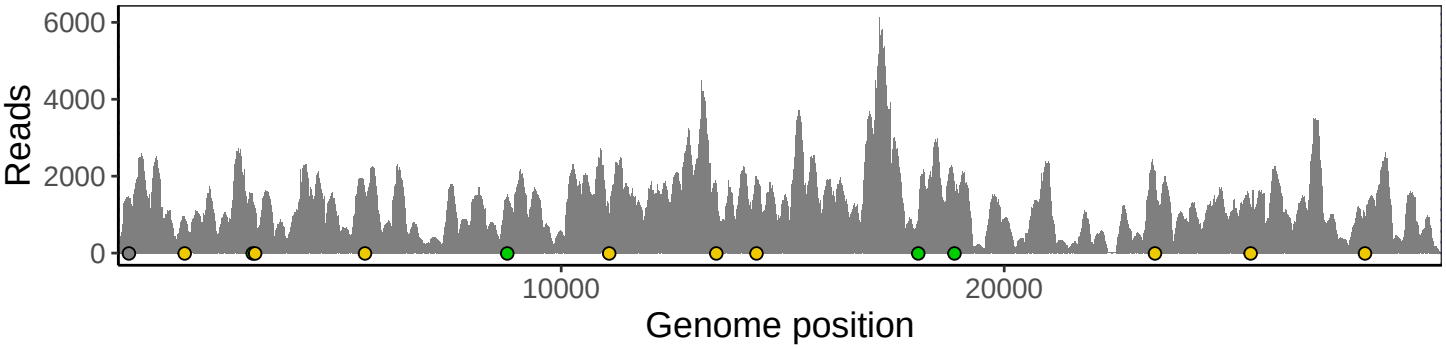


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

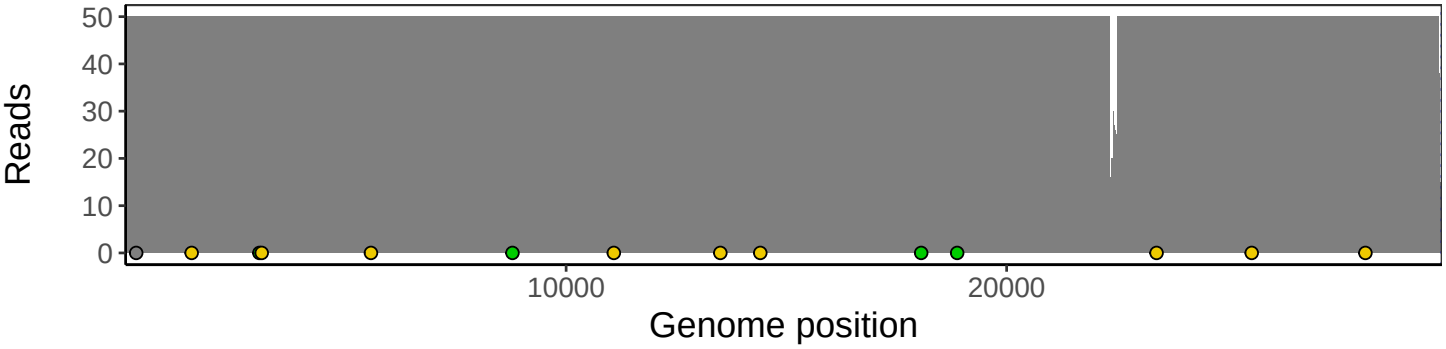


VSP0163-2 | 2020-05-22 | NP-OP | 266no-q | 7650000 genomes | single experiment

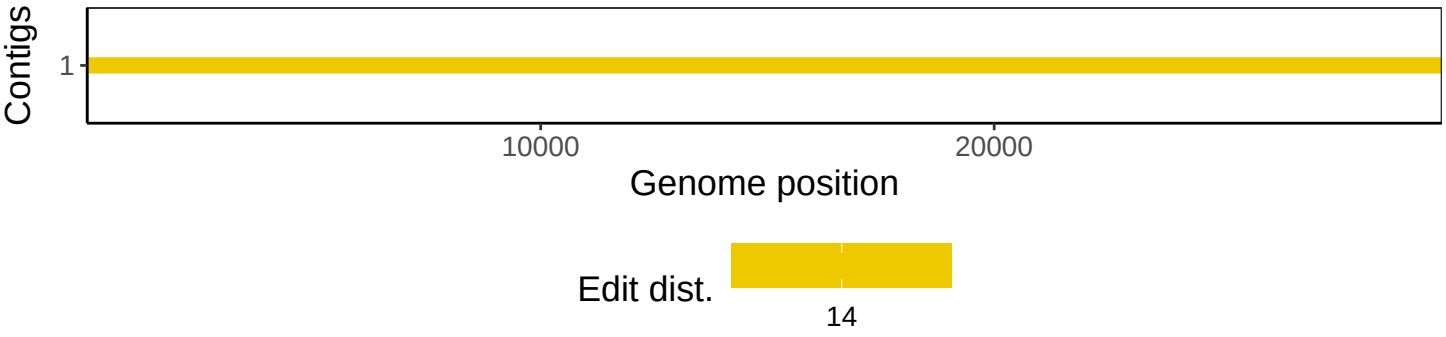
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

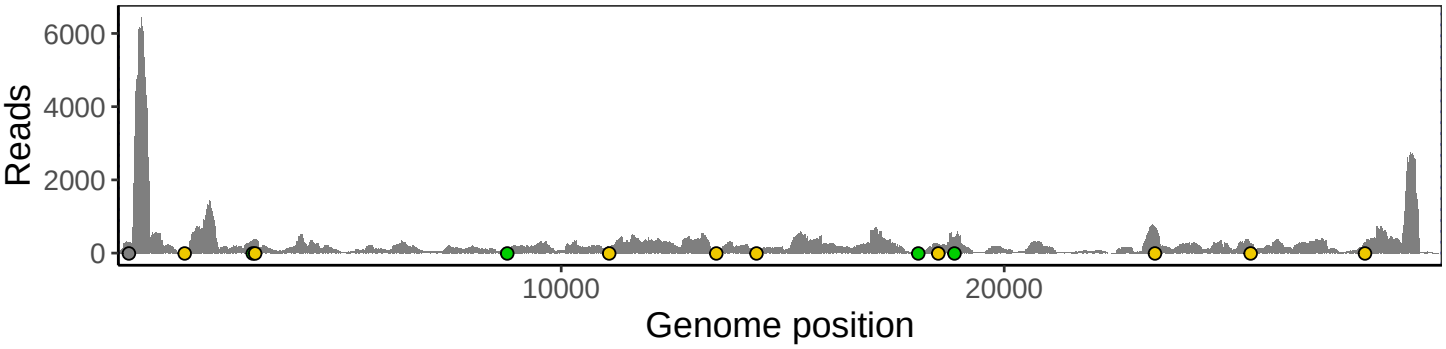


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

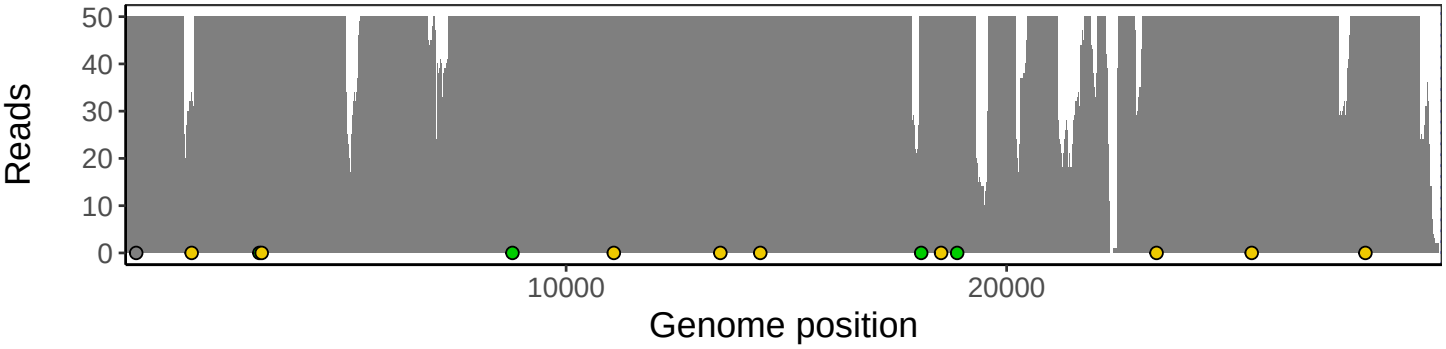


VSP0319-1 | 2020-05-20 | NP-OP | 266no-q | 143500 genomes | single experiment

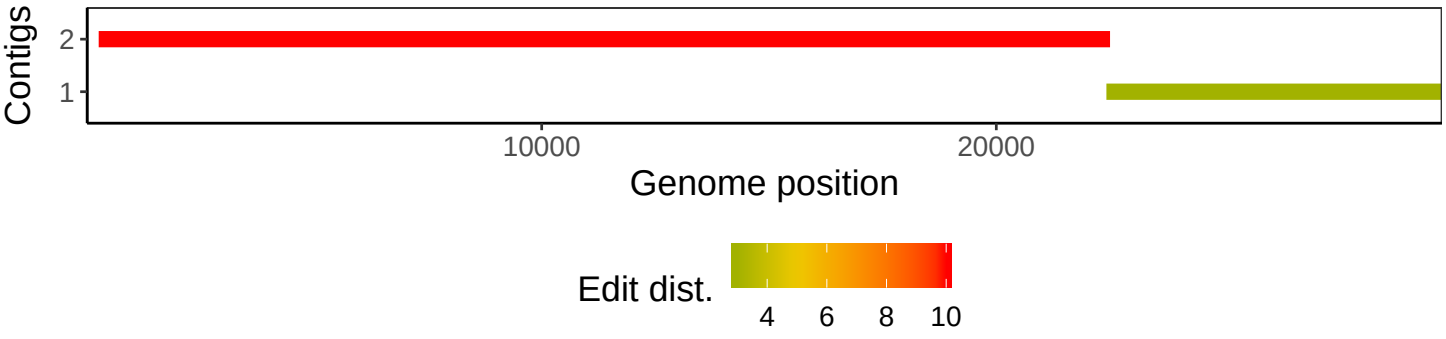
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

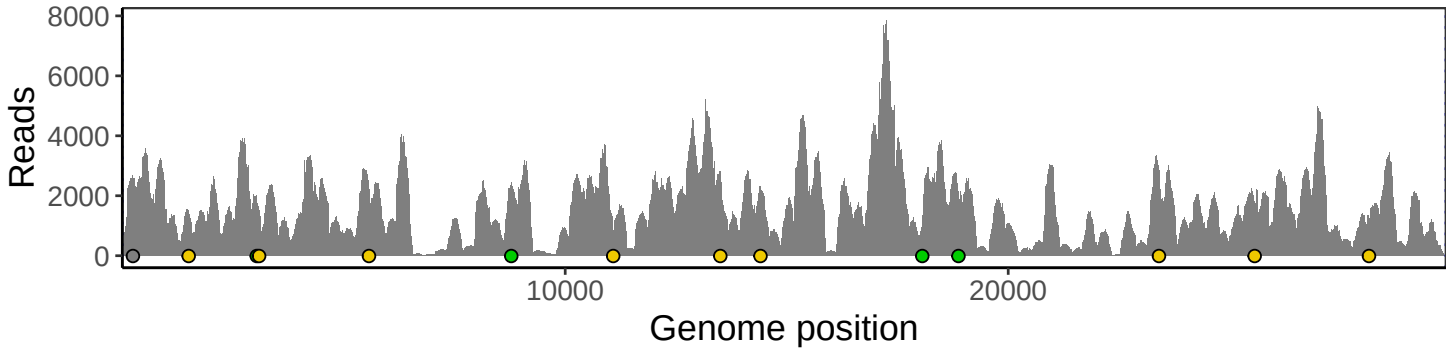


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

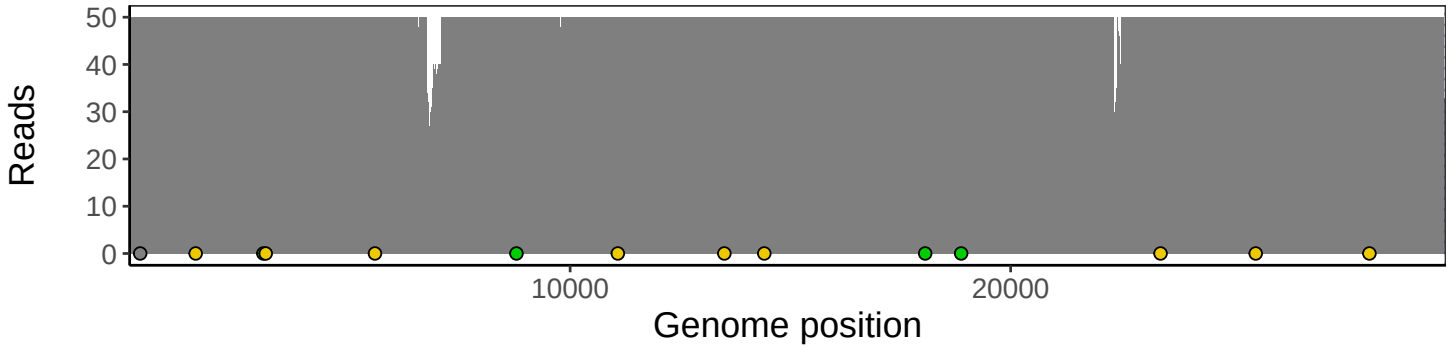


VSP0320-1 | 2020-05-22 | NP-OP | 266no-q | 5750000 genomes | single experiment

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.