COVID-19 subject deWit_RM7_Vehicle

2020-10-02

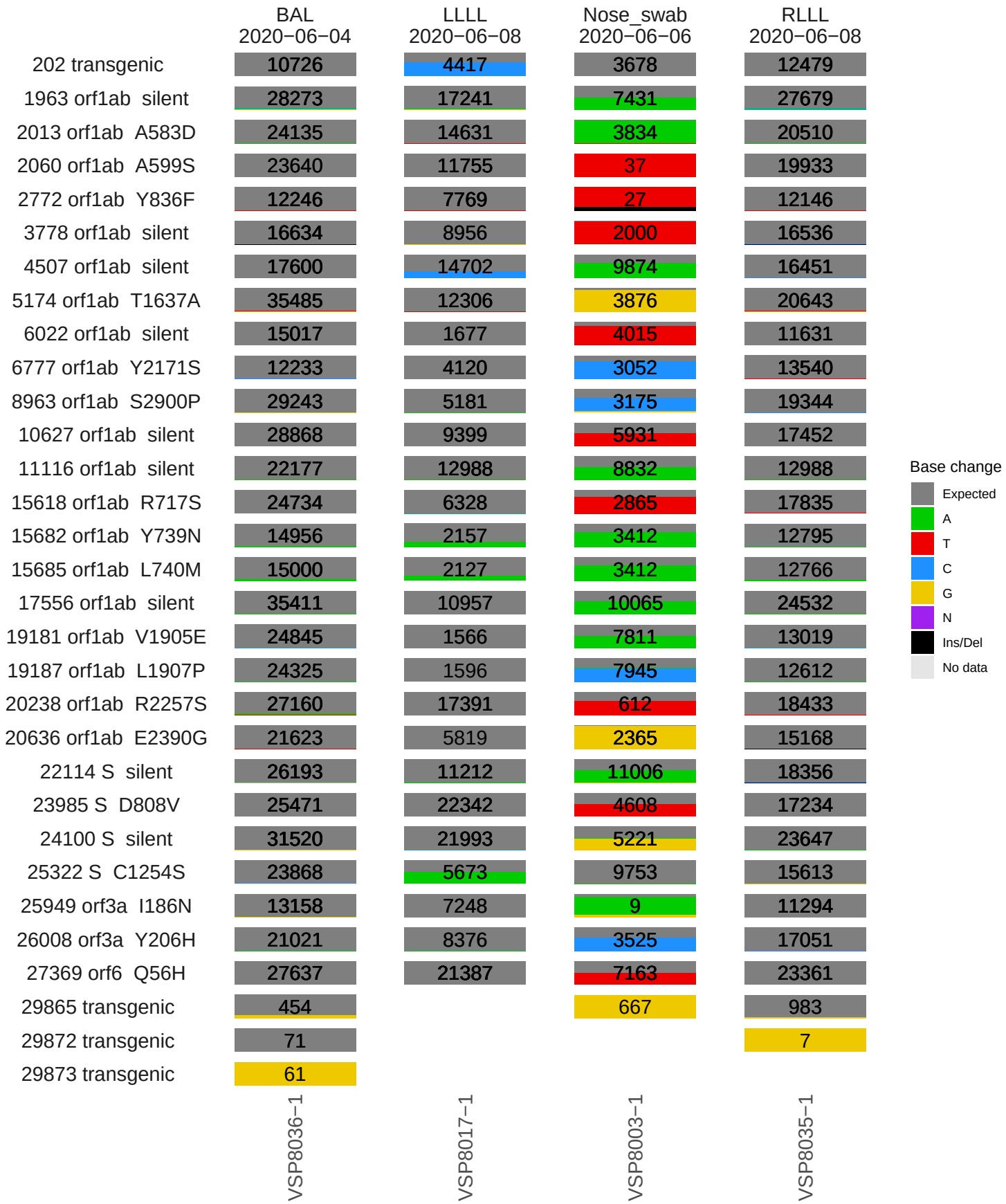
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. The code base for this analysis can be found (here).

Table 1. Sample summary.

Experiment	Type	Input genomes	Sample type	Sample date	Largest contig (KD)	Reference read coverage	Reference read coverage (>= 5 reads)
VSP8003-1	single experiment	NA	Nose_swab	2020-06-06	8.63	99.3%	99.0%
VSP8015-1	single experiment	NA	Rectal_swab	2020-06-03	29.98	100.0%	99.9%
VSP8017-1	single experiment	NA	LLLL	2020-06-08	11.71	99.9%	99.9%
VSP8035-1	single experiment	NA	RLLL	2020-06-08	29.84	100.0%	100.0%
VSP8036-1	single experiment	NA	BAL	2020-06-04	29.87	100.0%	100.0%

Variants shared across samples

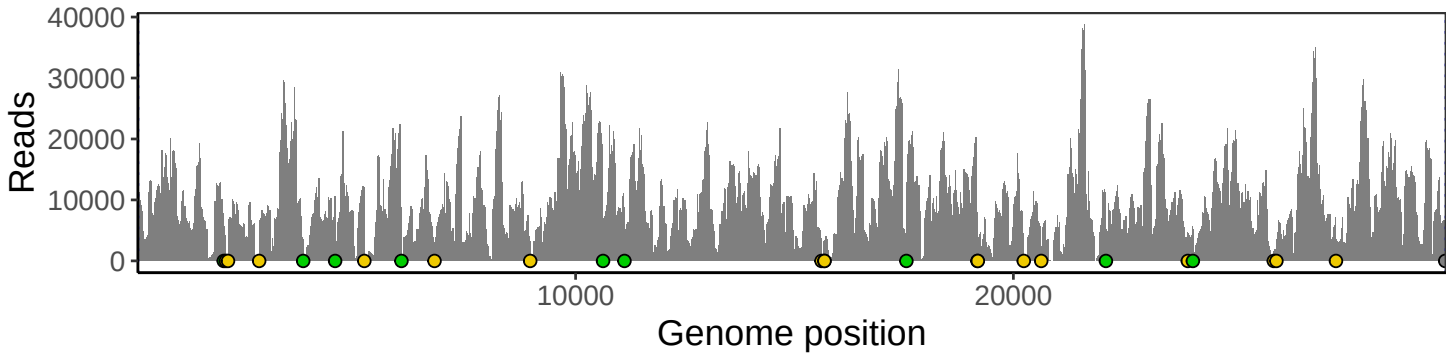
The heat map below shows how variants (reference genome USA-WA1-2020) are shared across subject samples where the percent variance is colored. Variants are called if a variant position is covered by 5 for more reads, the alternative base is found in $> 50\%$ of read pairs and the variant yields a PHRED 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.


BAL
2020−06−04
LLLL
2020−06−08
Nose_swab
2020−06−06
RLLL
2020−06−08
202 transgenic
1963 orf1ab silent
2013 orf1ab A583D
2060 orf1ab A599S
2772 orf1ab Y836F
3778 orf1ab silent
4507 orf1ab silent
5174 orf1ab T1637A
6022 orf1ab silent
6777 orf1ab Y2171S
8963 orf1ab S2900P
10627 orf1ab silent
11116 orf1ab silent
15618 orf1ab R717S
15682 orf1ab Y739N
15685 orf1ab L740M
17556 orf1ab silent
19181 orf1ab V1905E
19187 orf1ab L1907P
20238 orf1ab R2257S
20636 orf1ab E2390G
22114 S silent
23985 S D808V
24100 S silent
25322 S C1254S
25949 orf3a I186N
26008 orf3a Y206H
27369 orf6 Q56H
29865 transgenic
29872 transgenic
29873 transgenic
10726
28273
24135
23640
12246
16634
17600
35485
15017
12233
29243
28868
22177
24734
14956
15000
35411
24845
24325
27160
21623
26193
25471
31520
23868
13158
21021
27637
454
71
61
4417
17241
14631
11755
7769
8956
14702
12306
1677
4120
5181
9399
12988
6328
2157
2127
10957
1566
1596
17391
5819
11212
22342
21993
5673
7248
8376
21387
3678
7431
3834
37
27
2000
9874
3876
4015
3052
3175
5931
8832
2865
3412
3412
10065
7811
7945
612
2365
11006
4608
5221
9753
9
3525
7163
667
12479
27679
20510
19933
12146
16536
16451
20643
11631
13540
19344
17452
12988
17835
12795
12766
24532
13019
12612
18433
15168
18356
17234
23647
15613
11294
17051
23361
983
7
Base change
Expected
A
T
C
G
N
Ins/Del
No data
VSP8036−1
VSP8017−1
VSP8003−1
VSP8035−1

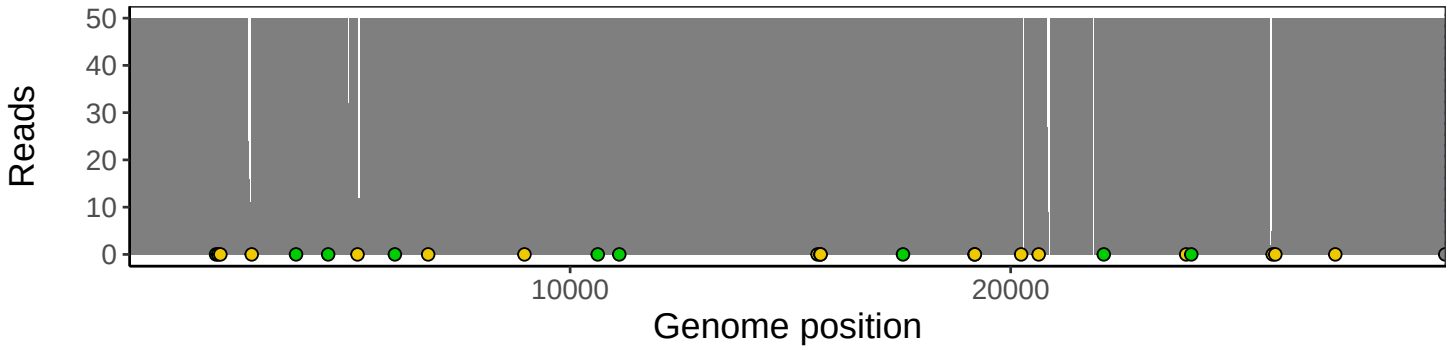
Analyses of individual experiments and composite results.

VSP8003-1 | 2020-06-06 | Nose_swab | SRR11783603 | genomes | single experiment

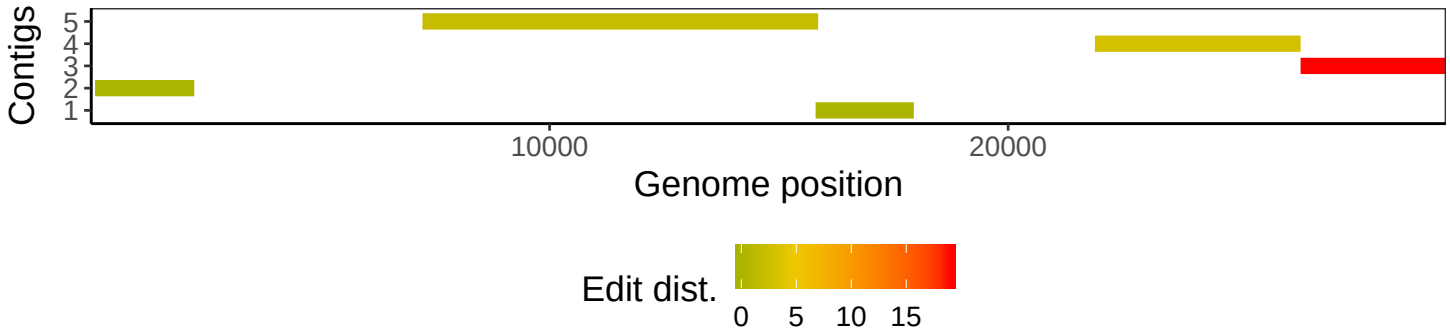
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

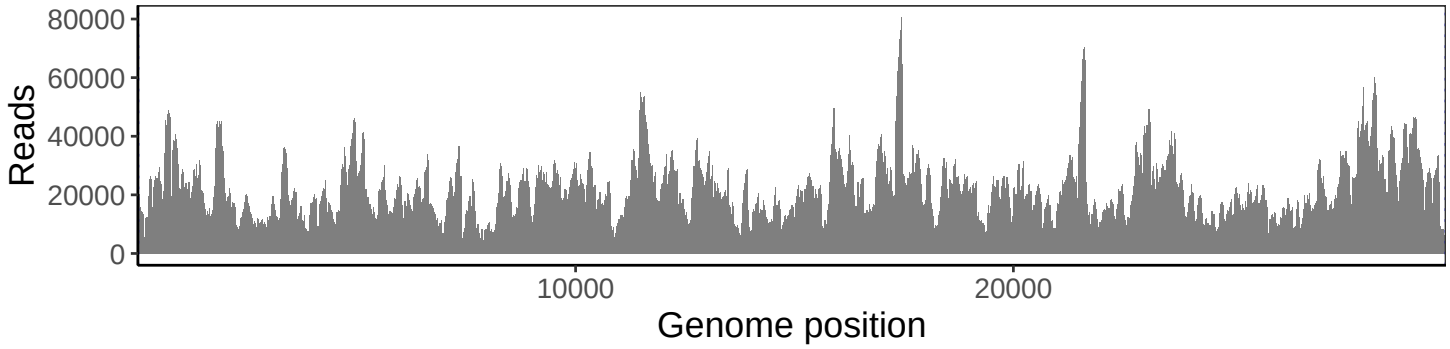


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

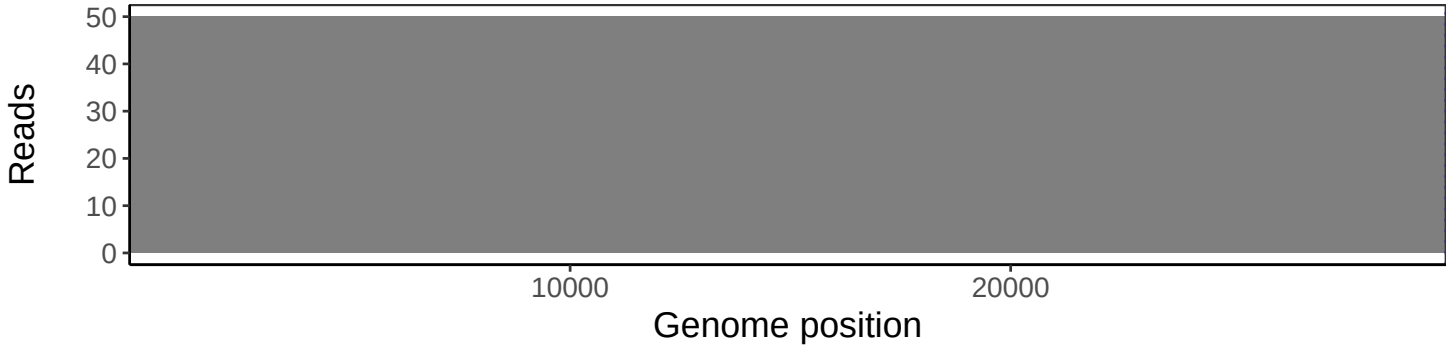


VSP8015-1 | 2020-06-03 | Rectal_swab | SRR11783600 | genomes | single experiment

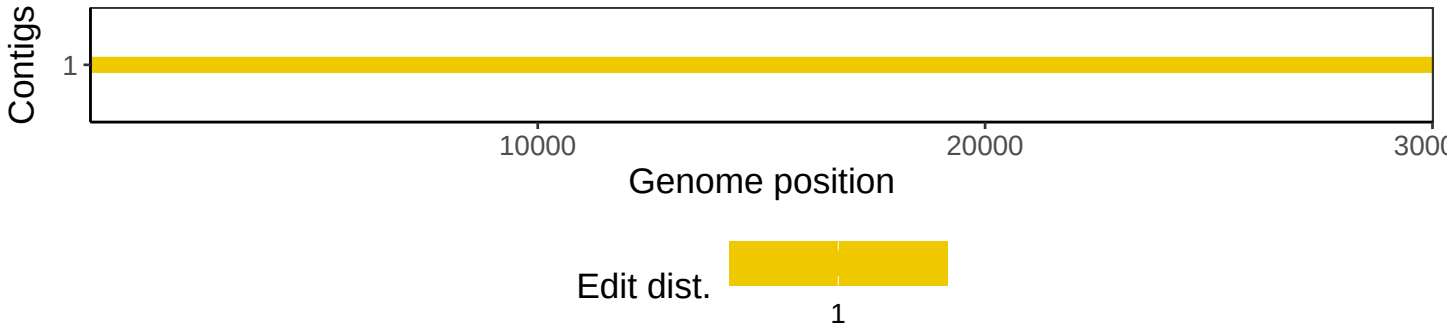
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

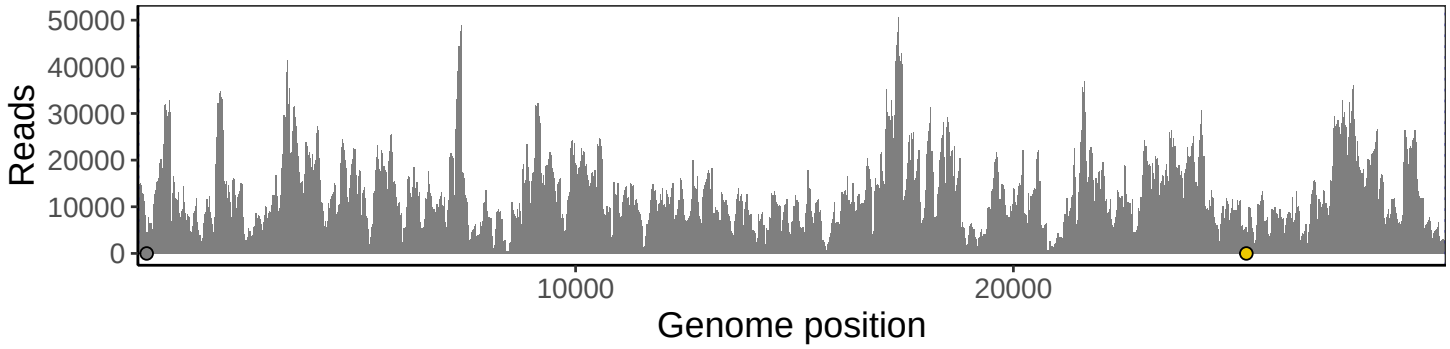


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

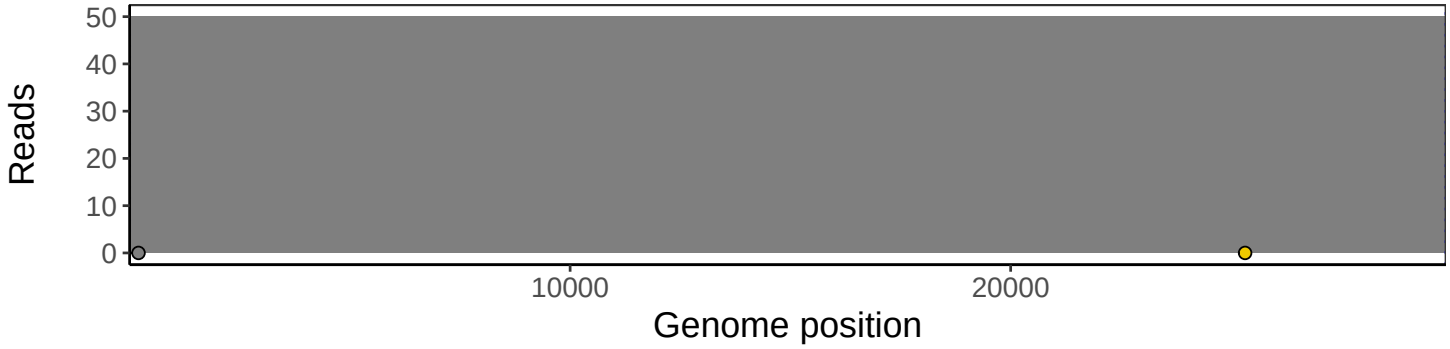


VSP8017-1 | 2020-06-08 | LLLL | SRR11783604 | genomes | single experiment

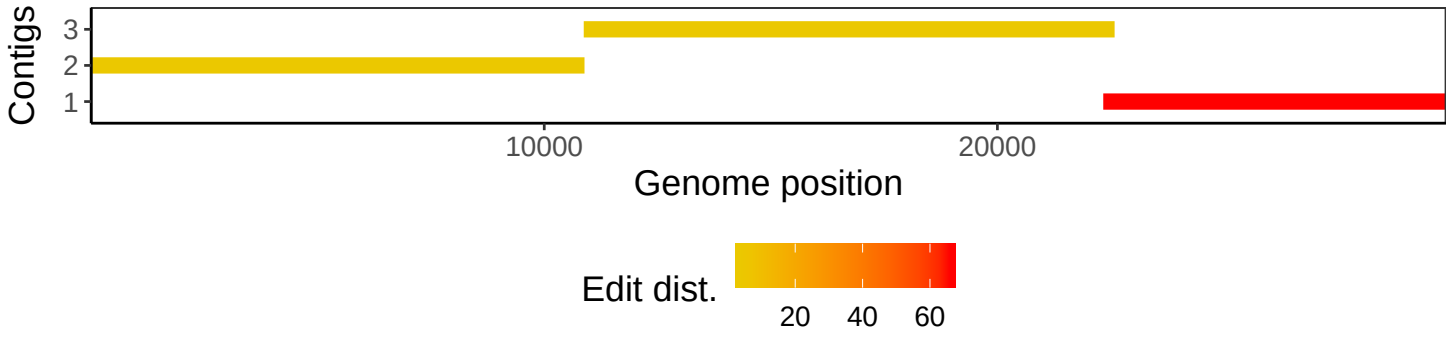
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

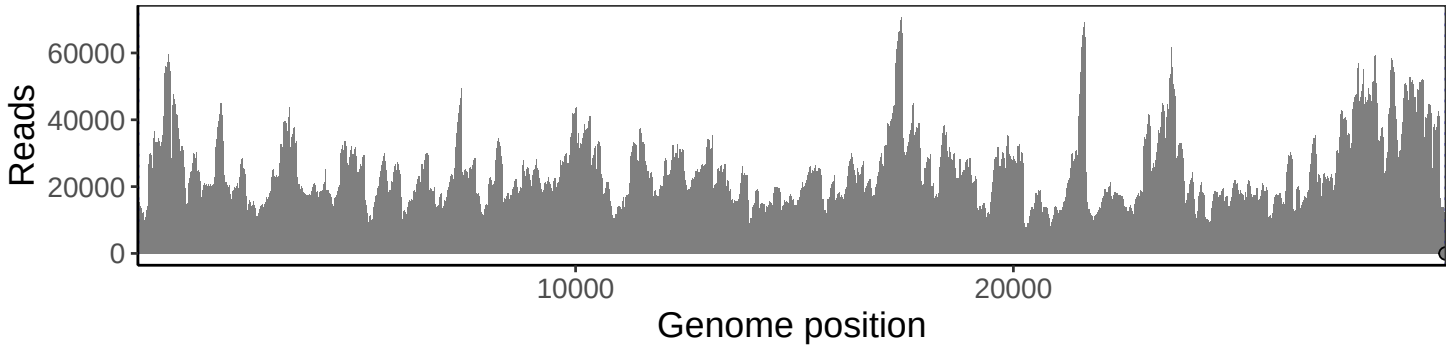


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

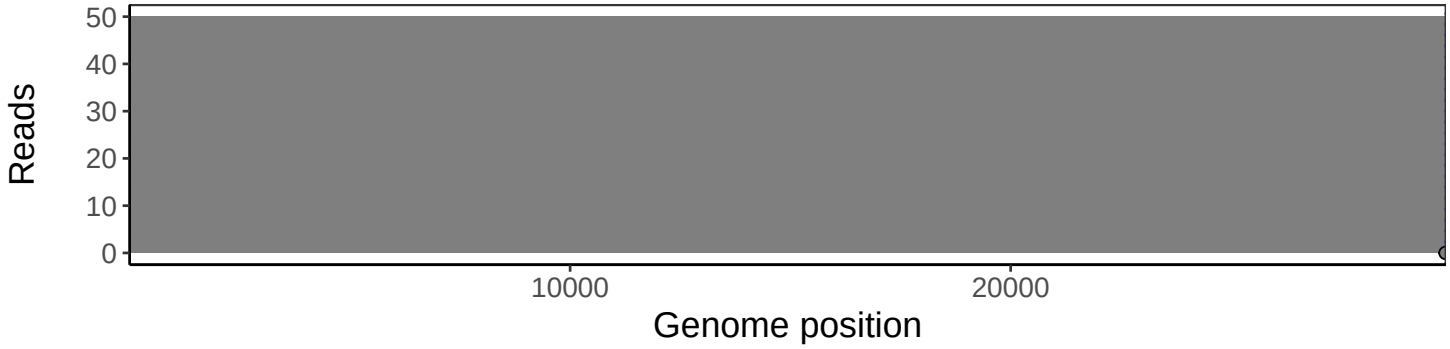


VSP8035-1 | 2020-06-08 | RLLL | SRR11783601 | genomes | single experiment

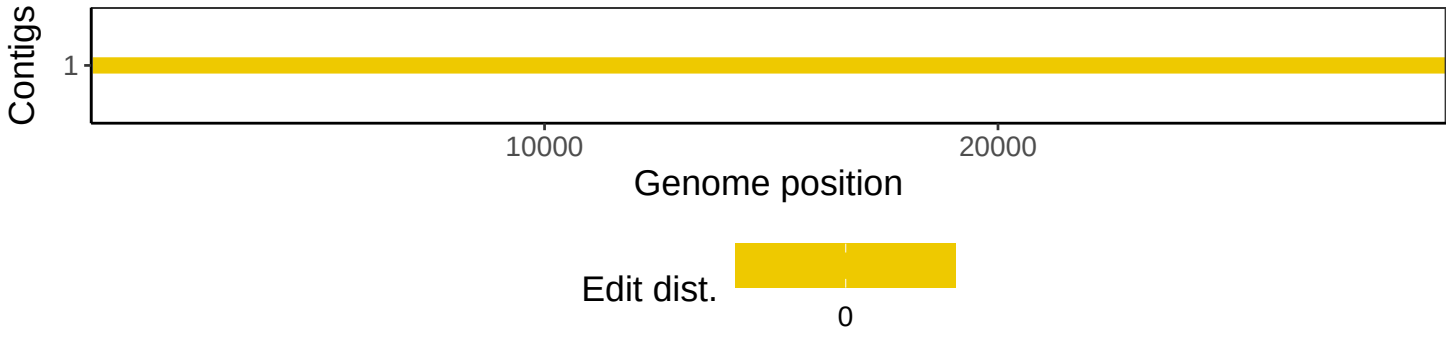
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

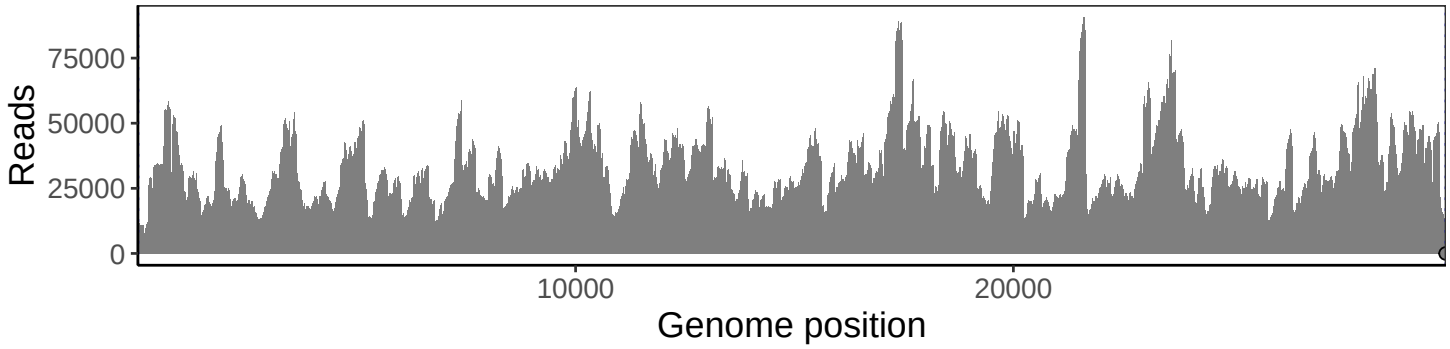


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

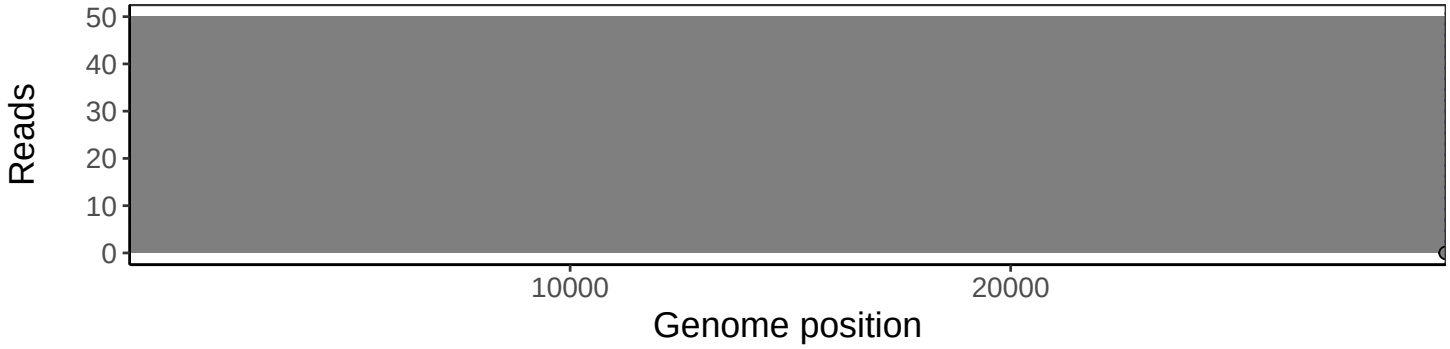


VSP8036-1 | 2020-06-04 | BAL | SRR11783605 | genomes | single experiment

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.