COVID-19 subject 290

2021-01-31

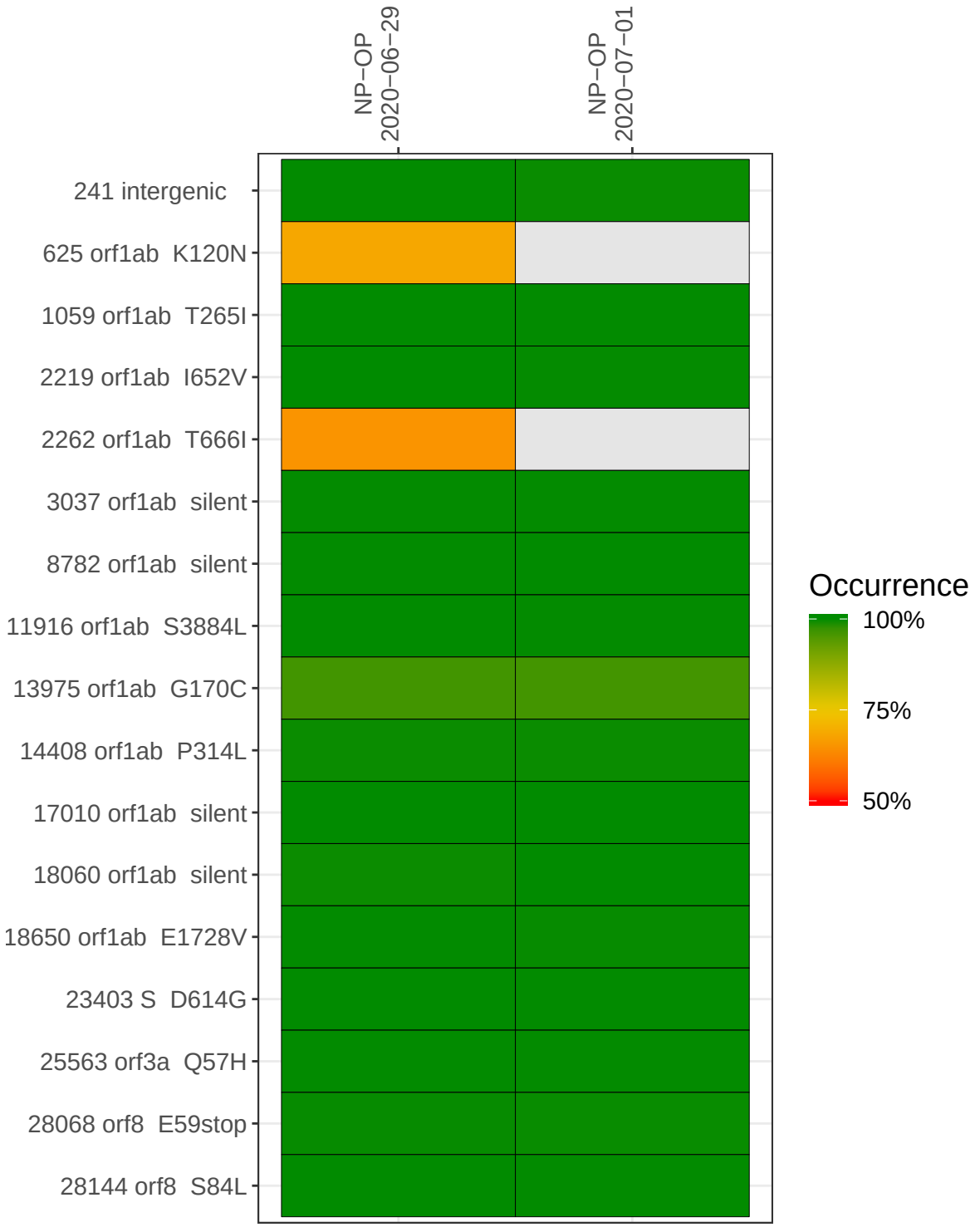
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with > 90% sequence coverage.

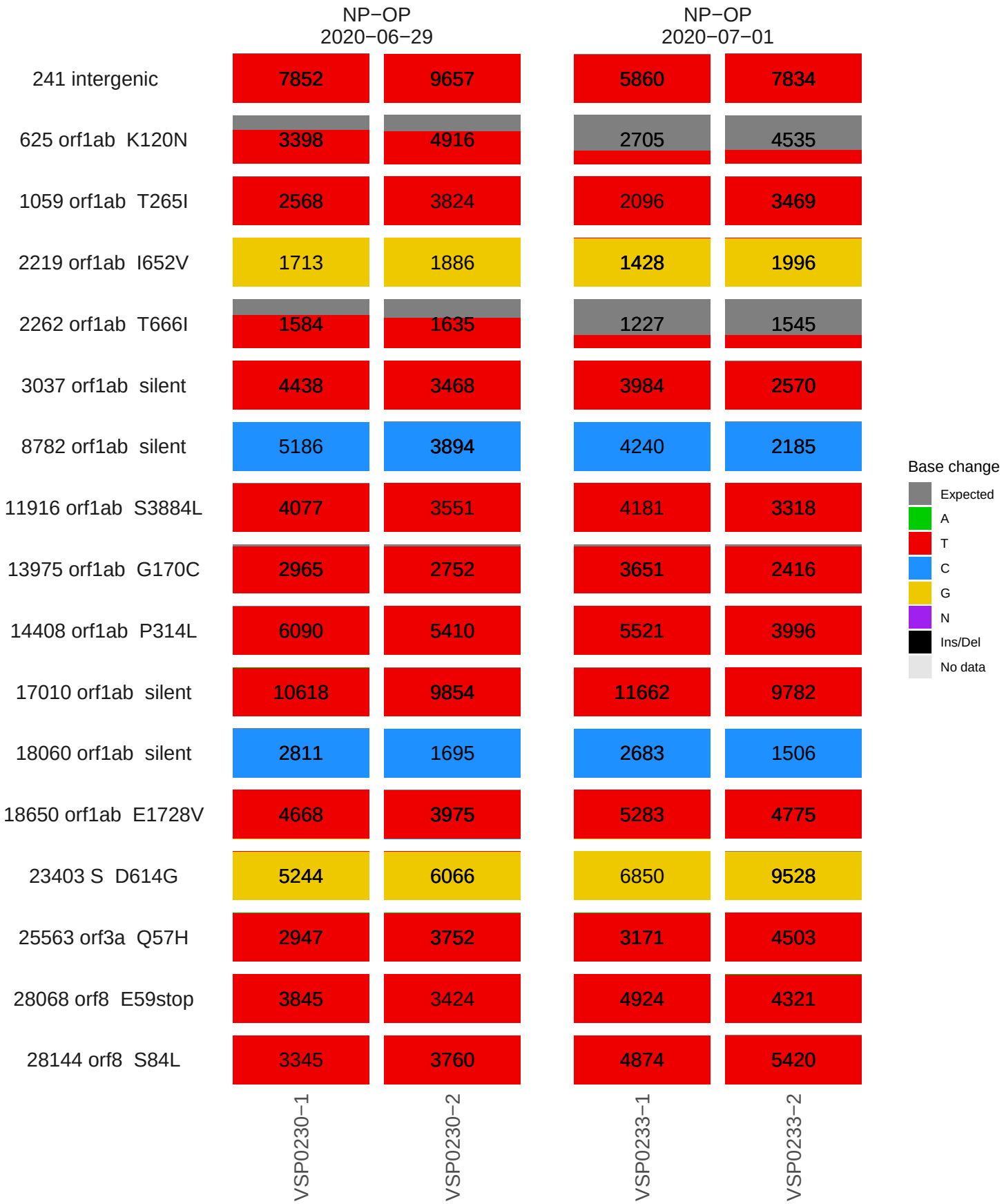
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0230	composite	NA	NP-OP	2020-06-29	29.80	B.1.3.1	99.8%	99.8%
VSP0233	composite	NA	NP-OP	2020-07-01	24.18	B.1.3.1	99.8%	99.8%
VSP0230-1	single experiment	11750	NP-OP	2020-06-29	27.24	B.1.3.1	99.8%	99.8%
VSP0230-2	single experiment	11750	NP-OP	2020-06-29	29.80	B.1.3.1	99.8%	99.8%
VSP0233-1	single experiment	27900	NP-OP	2020-07-01	29.86	B.1.3.1	99.8%	99.8%
VSP0233-2	single experiment	27900	NP-OP	2020-07-01	28.54	B.1.3.1	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 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.

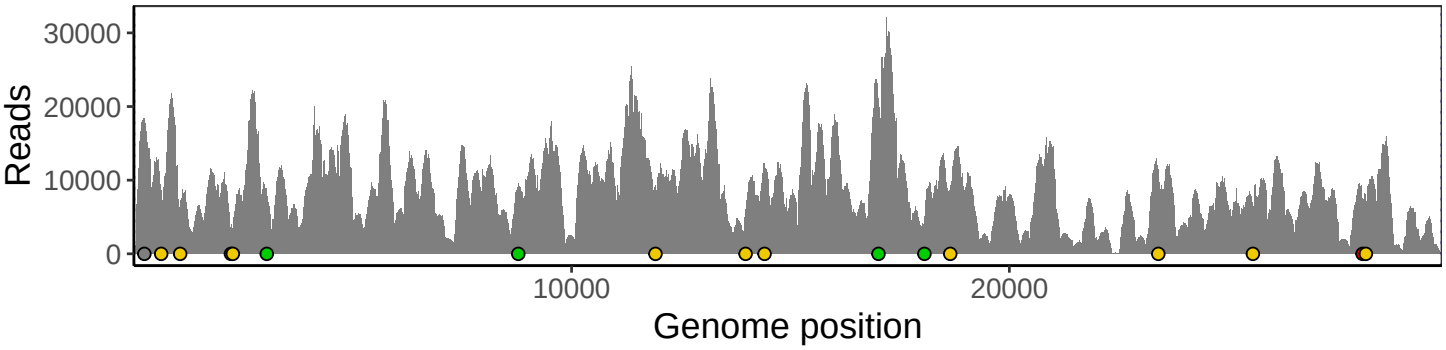



NP−OP
2020−06−29
NP−OP
2020−07−01
241 intergenic
625 orf1ab K120N
1059 orf1ab T265I
2219 orf1ab I652V
2262 orf1ab T666I
3037 orf1ab silent
8782 orf1ab silent
11916 orf1ab S3884L
13975 orf1ab G170C
14408 orf1ab P314L
17010 orf1ab silent
18060 orf1ab silent
18650 orf1ab E1728V
23403 S D614G
25563 orf3a Q57H
28068 orf8 E59stop
28144 orf8 S84L
7852
9657
5860
7834
3398
4916
2705
4535
2568
3824
2096
3469
1713
1886
1428
1996
1584
1635
1227
1545
4438
3468
3984
2570
5186
3894
4240
2185
4077
3551
4181
3318
2965
2752
3651
2416
6090
5410
5521
3996
10618
9854
11662
9782
2811
1695
2683
1506
4668
3975
5283
4775
5244
6066
6850
9528
2947
3752
3171
4503
3845
3424
4924
4321
3345
3760
4874
5420
VSP0230−1
VSP0230−2
VSP0233−1
VSP0233−2
Base change
Expected
A
T
C
G
N
Ins/Del
No data

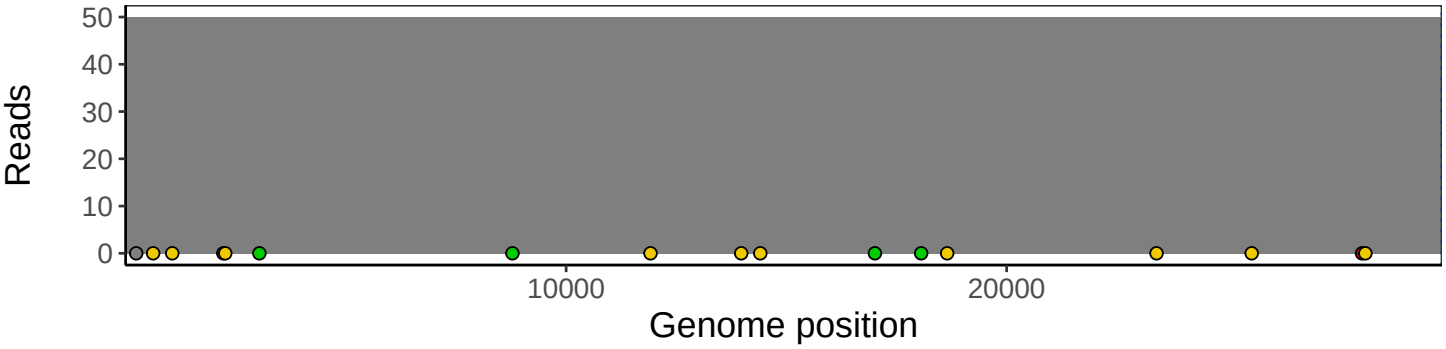
Analyses of individual experiments and composite results

VSP0230 | 2020-06-29 | NP-OP | 290no-q | composite result

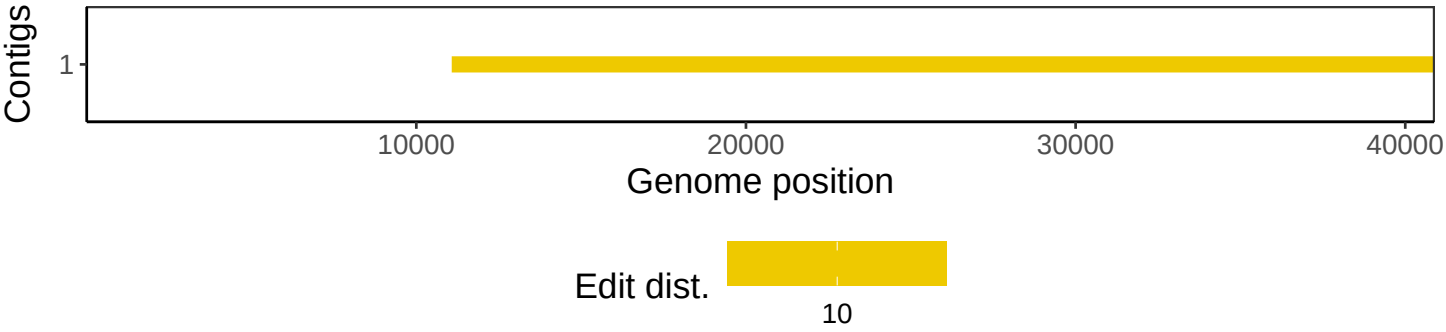
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

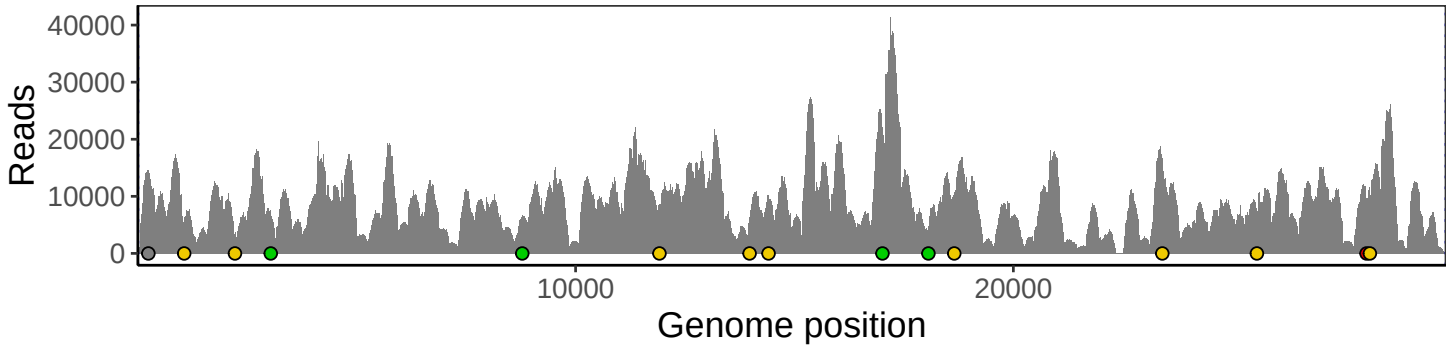


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

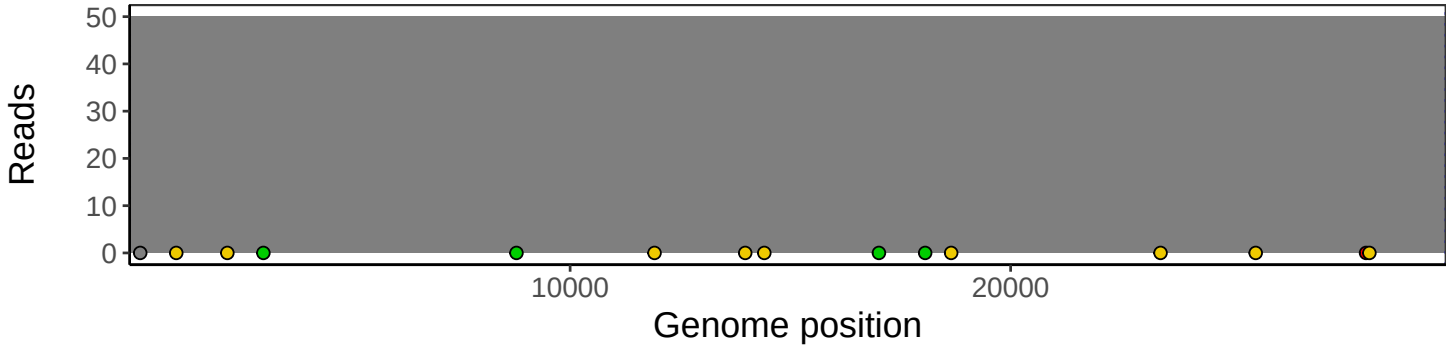


VSP0233 | 2020-07-01 | NP-OP | 290no-q | composite result

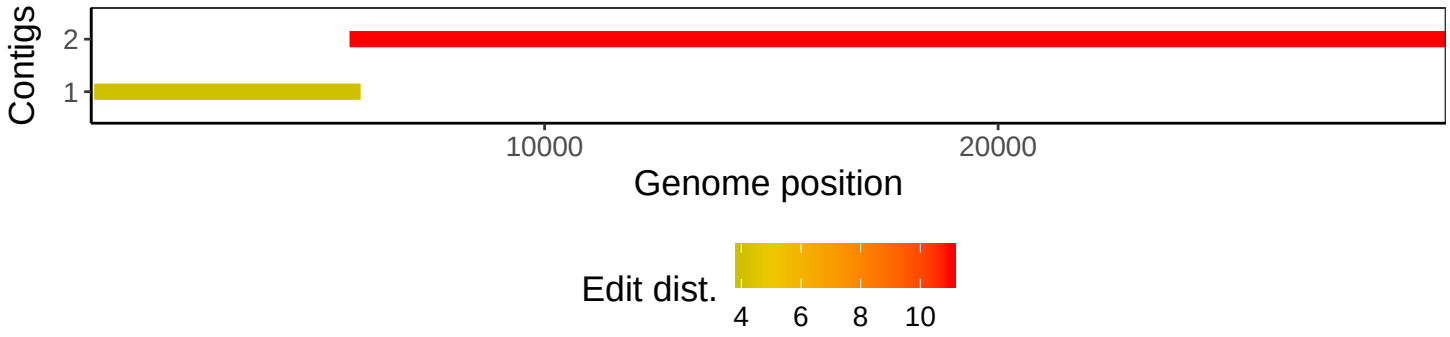
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

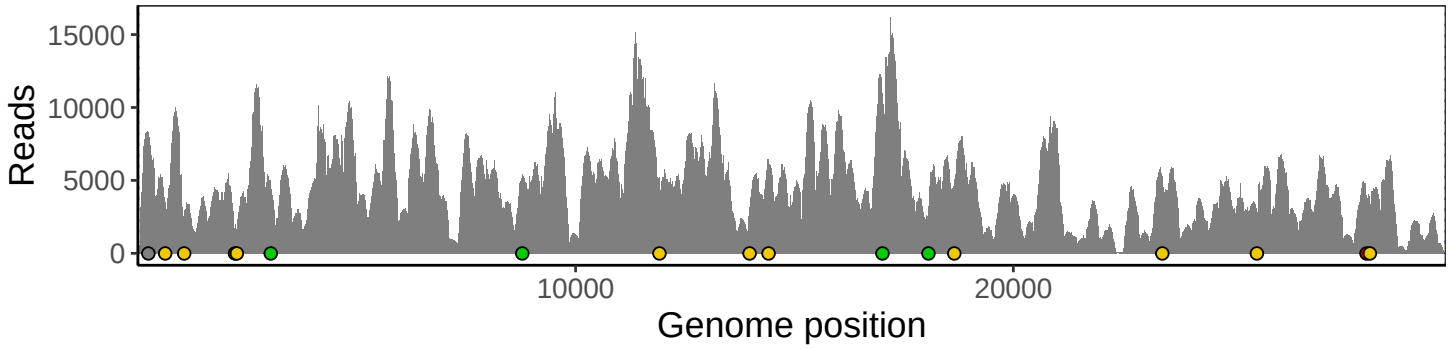


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

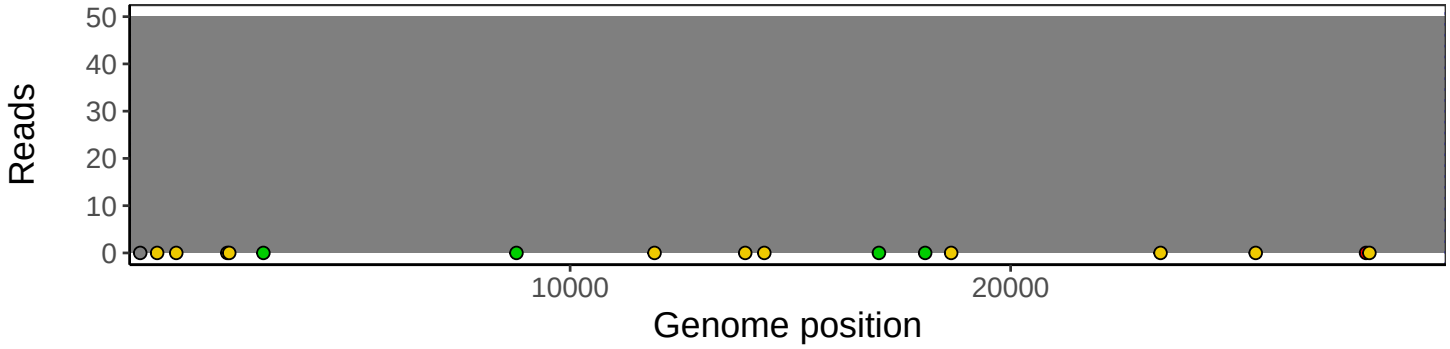


VSP0230-1 | 2020-06-29 | NP-OP | 290no-q | 11750 genomes | single experiment

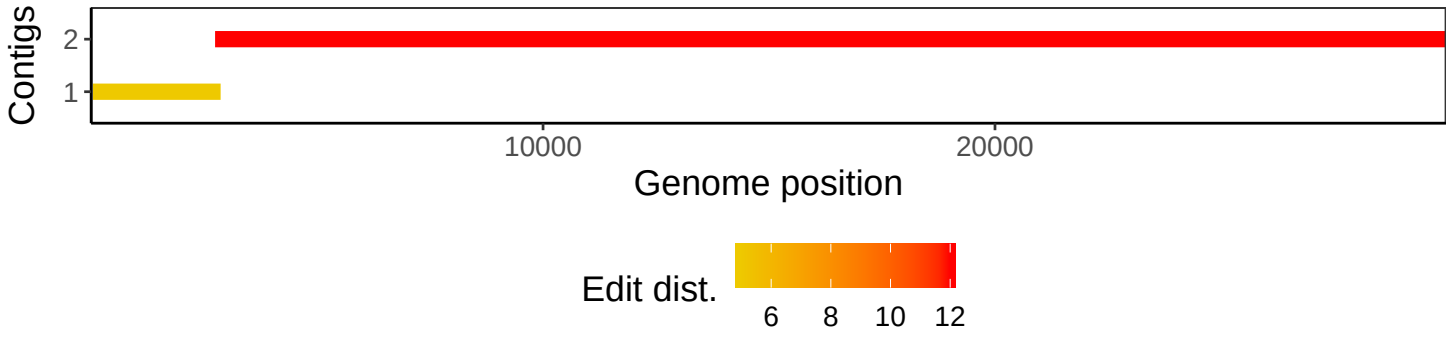
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

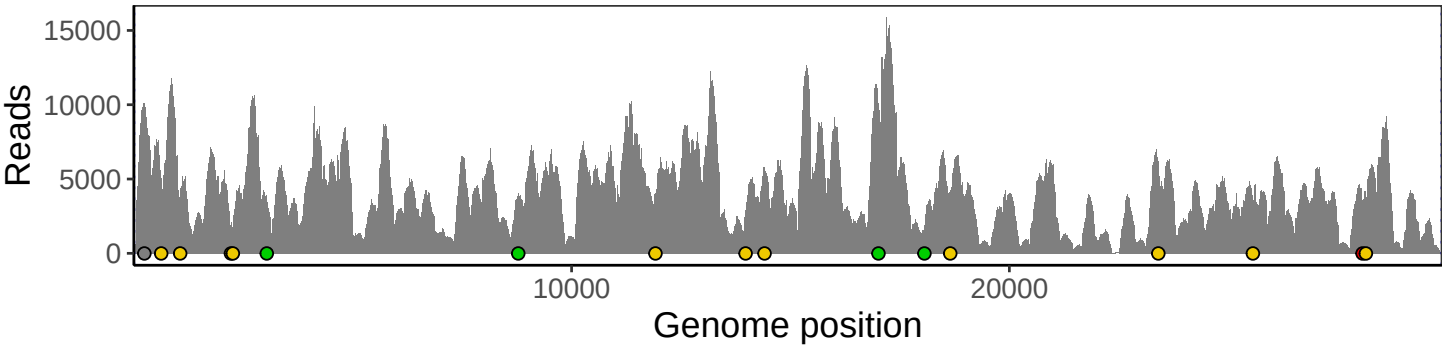


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

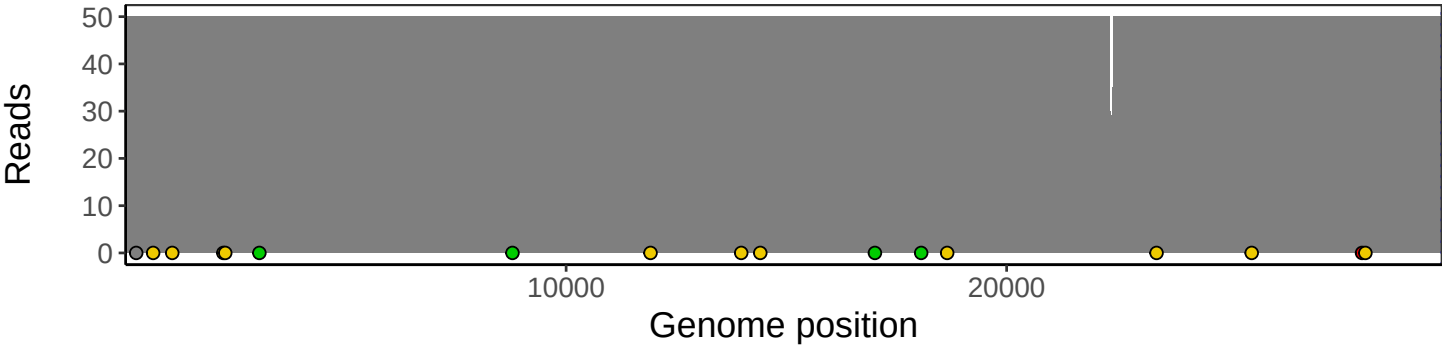


VSP0230-2 | 2020-06-29 | NP-OP | 290no-q | 11750 genomes | single experiment

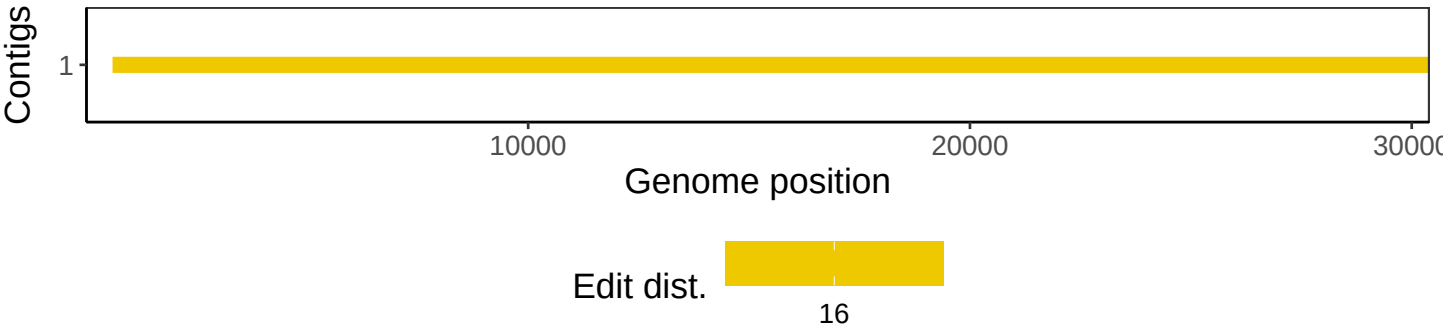
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

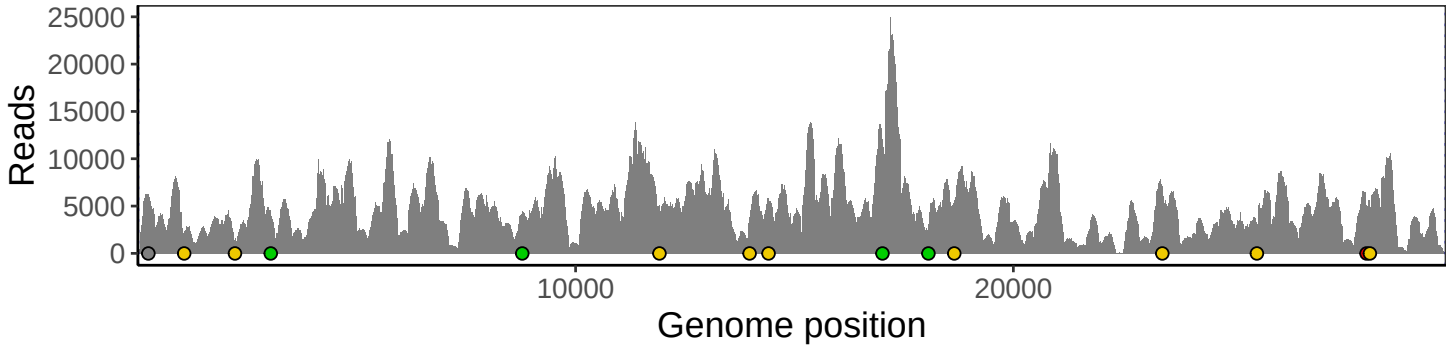


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

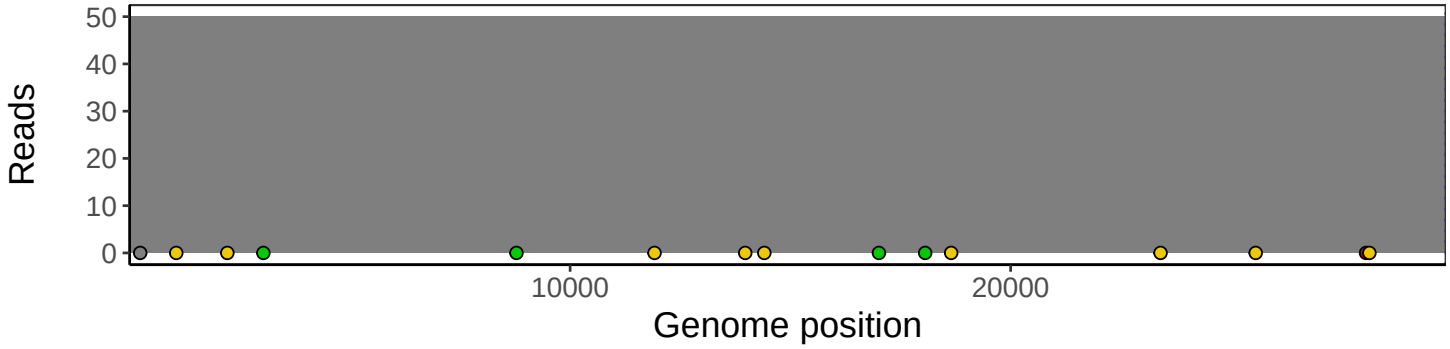


VSP0233-1 | 2020-07-01 | NP-OP | 290no-q | 27900 genomes | single experiment

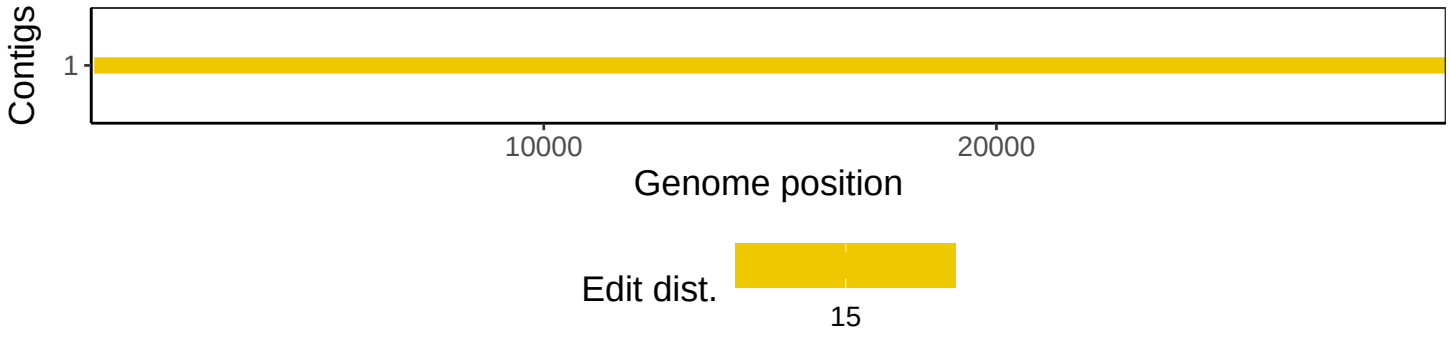
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

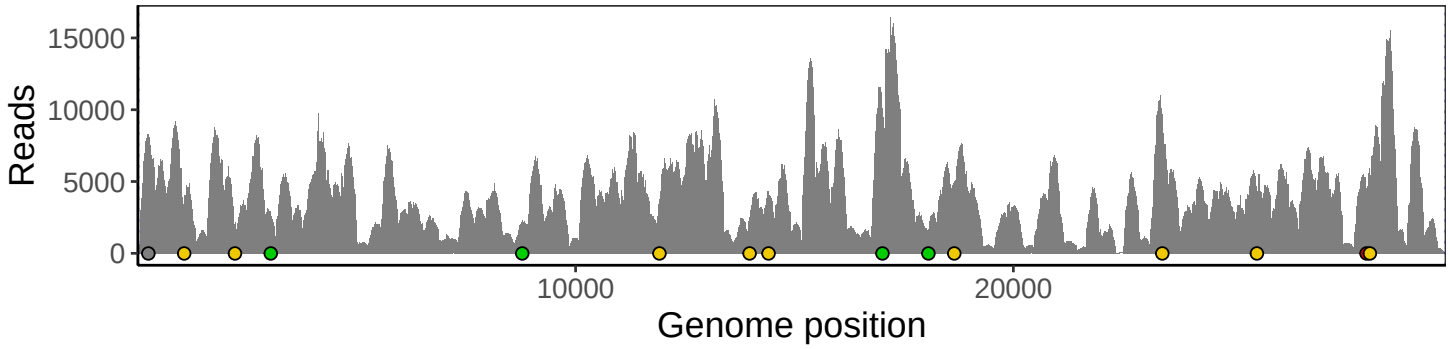


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

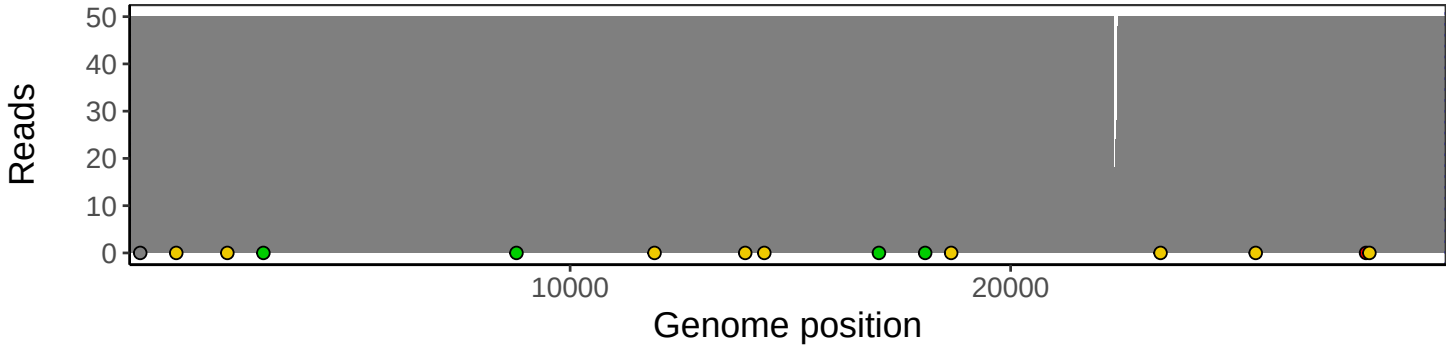


VSP0233-2 | 2020-07-01 | NP-OP | 290no-q | 27900 genomes | single experiment

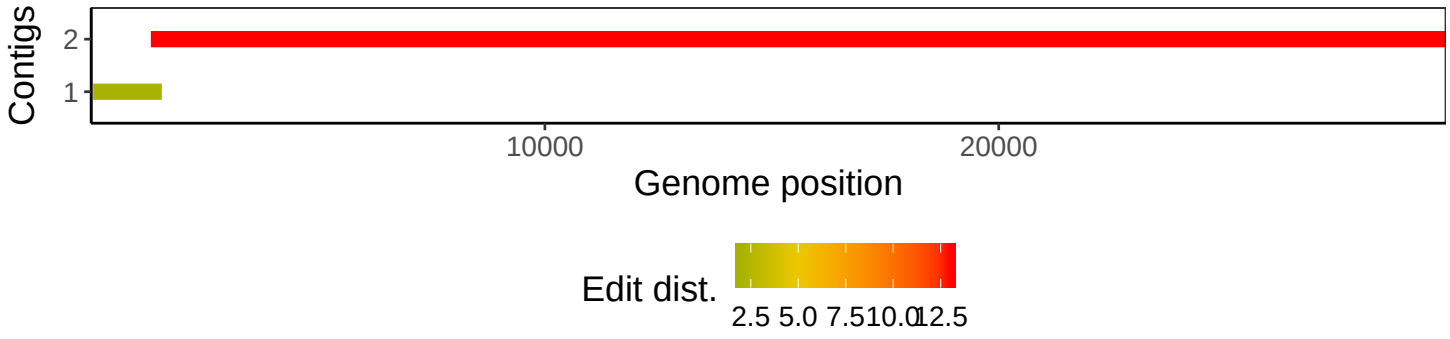
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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