COVID-19 subject 219

2021-01-27

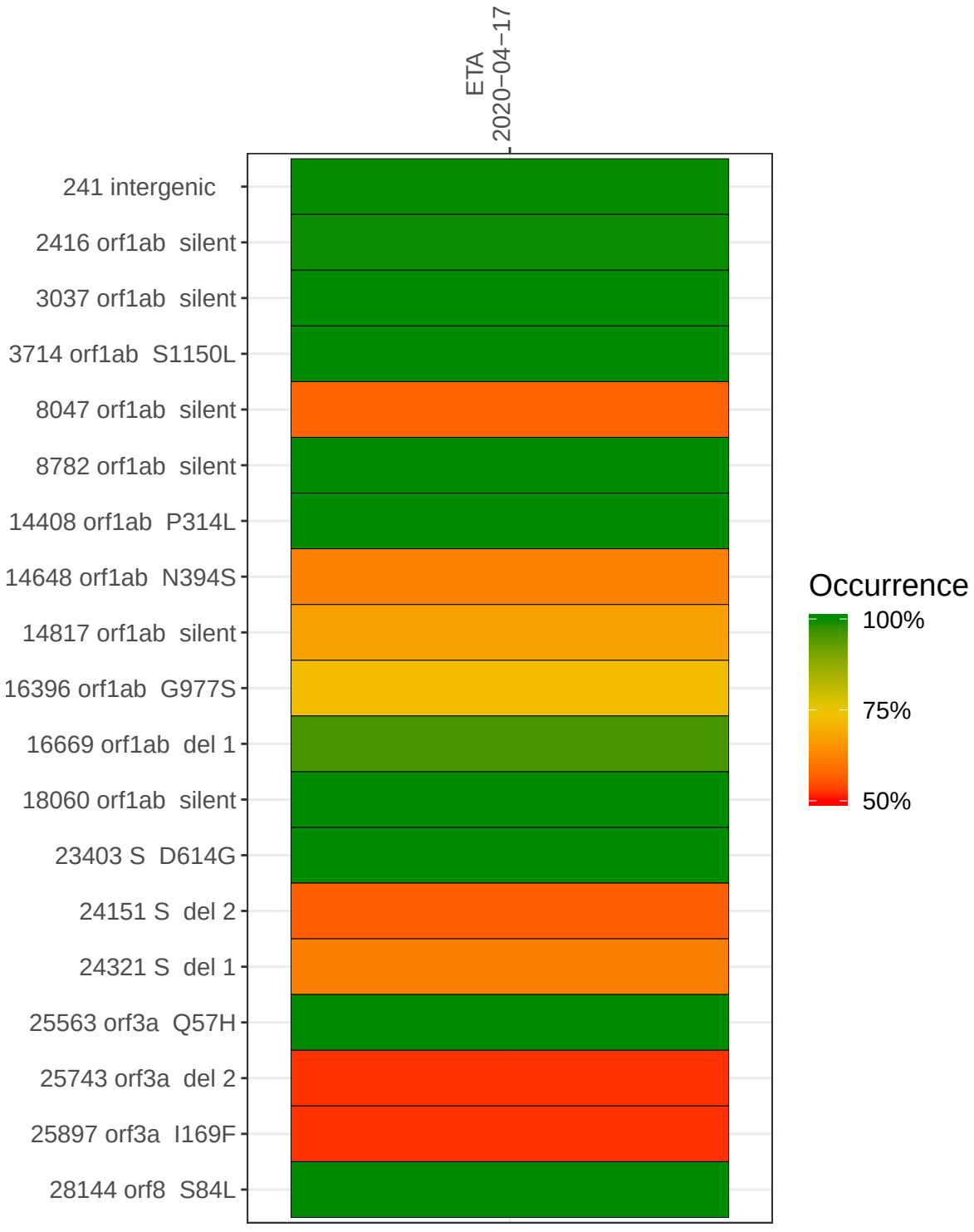
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

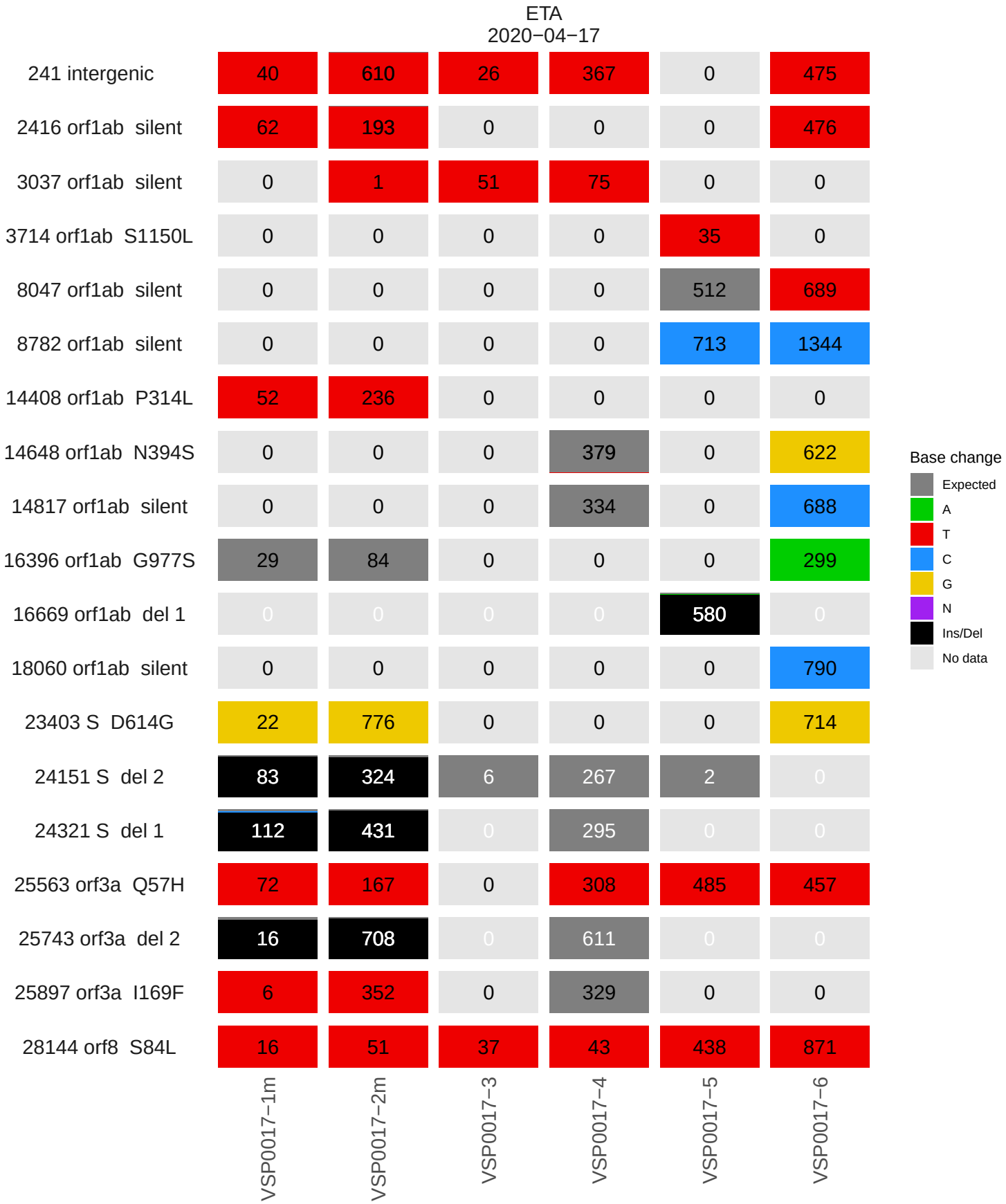
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0017	composite	NA	ETA	2020-04-17	7.66	B.1	93.3%	92.2%
VSP0017-1m	single experiment	NA	ETA	2020-04-17	0.93	NA	39.4%	35.1%
VSP0017-2m	single experiment	NA	ETA	2020-04-17	1.23	NA	47.3%	43.6%
VSP0017-3	single experiment	9350	ETA	2020-04-17	1.53	NA	48.8%	43.0%
VSP0017-4	single experiment	1870	ETA	2020-04-17	1.06	NA	49.2%	47.2%
VSP0017-5	single experiment	1870	ETA	2020-04-17	1.69	NA	44.0%	39.7%
VSP0017-6	single experiment	1870	ETA	2020-04-17	3.40	NA	64.2%	62.2%

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.

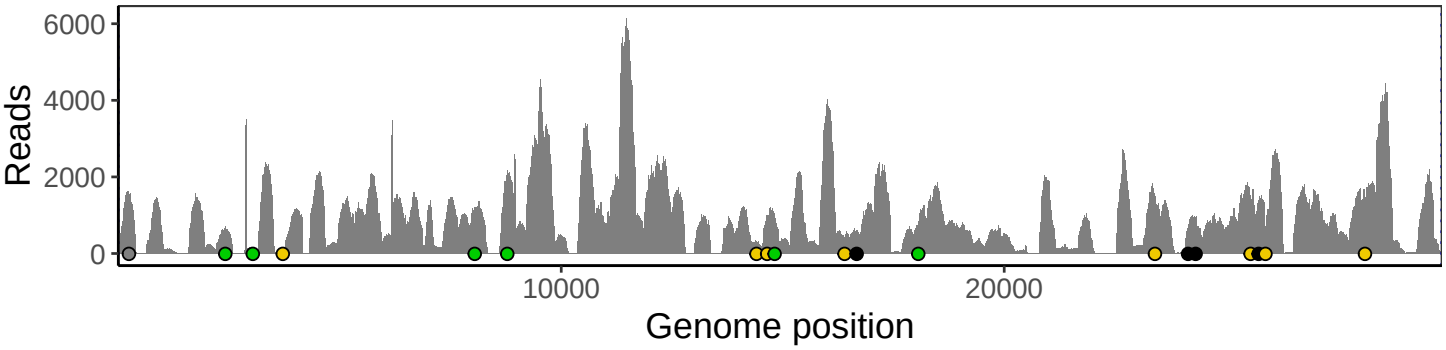



ETA
2020−04−17
241 intergenic
2416 orf1ab silent
3037 orf1ab silent
3714 orf1ab S1150L
8047 orf1ab silent
8782 orf1ab silent
14408 orf1ab P314L
14648 orf1ab N394S
14817 orf1ab silent
16396 orf1ab G977S
16669 orf1ab del 1
18060 orf1ab silent
23403 S D614G
24151 S del 2
24321 S del 1
25563 orf3a Q57H
25743 orf3a del 2
25897 orf3a I169F
28144 orf8 S84L
40 610 26 367 0 475
62 193 0 0 0 476
0 1 51 75 0 0
0 0 0 0 35 0
0 0 0 0 512 689
0 0 0 0 713 1344
52 236 0 0 0 0
0 0 0 379 0 622
0 0 0 334 0 688
29 84 0 0 0 299
0 0 0 0 580 0
0 0 0 0 0 790
22 776 0 0 0 714
83 324 6 267 2 0
112 431 0 295 0 0
72 167 0 308 485 457
16 708 0 611 0 0
6 352 0 329 0 0
16 51 37 43 438 871
VSP0017−1m
VSP0017−2m
VSP0017−3
VSP0017−4
VSP0017−5
VSP0017−6
Base change
Expected
A
T
C
G
N
Ins/Del
No data

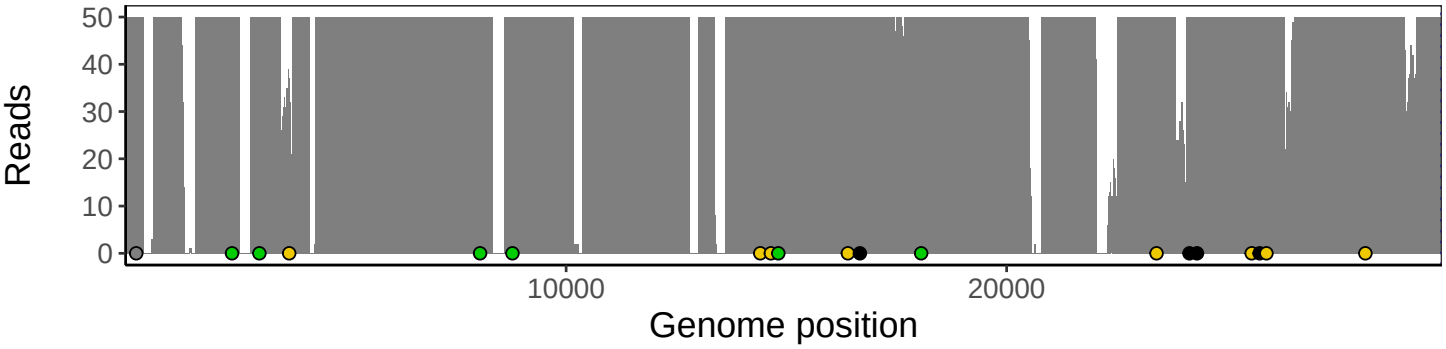
Analyses of individual experiments and composite results

VSP0017 | 2020-04-17 | ETA | 219-tri | 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.



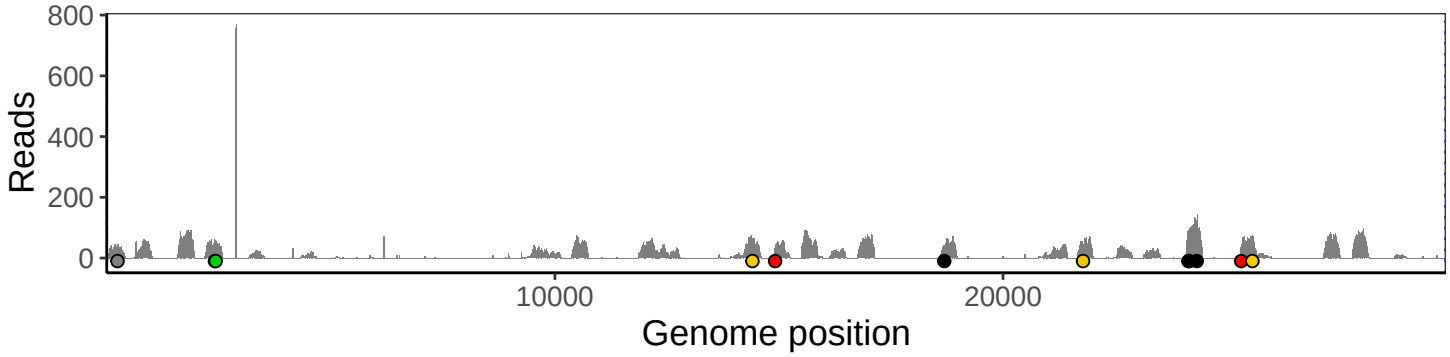
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



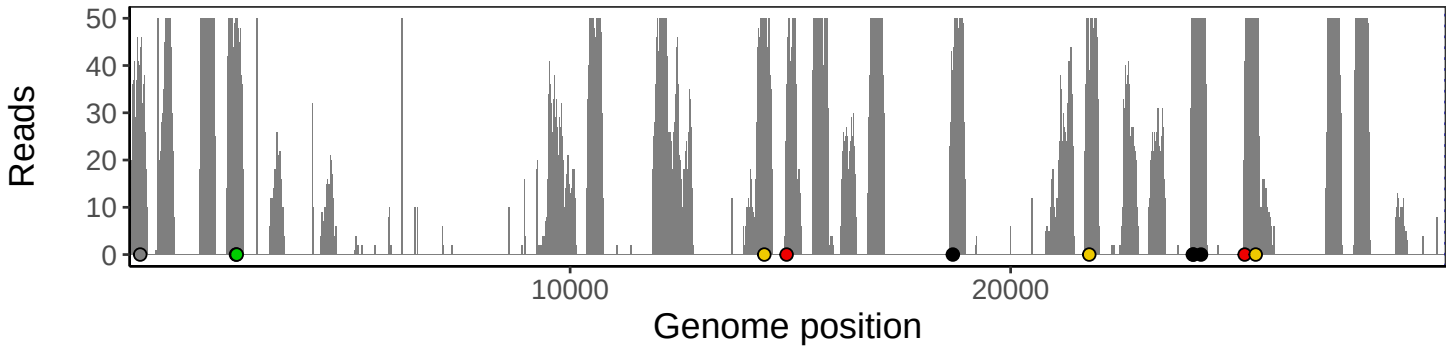
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

VSP0017-1m | 2020-04-17 | ETA | 219-tri | genomes | single experiment

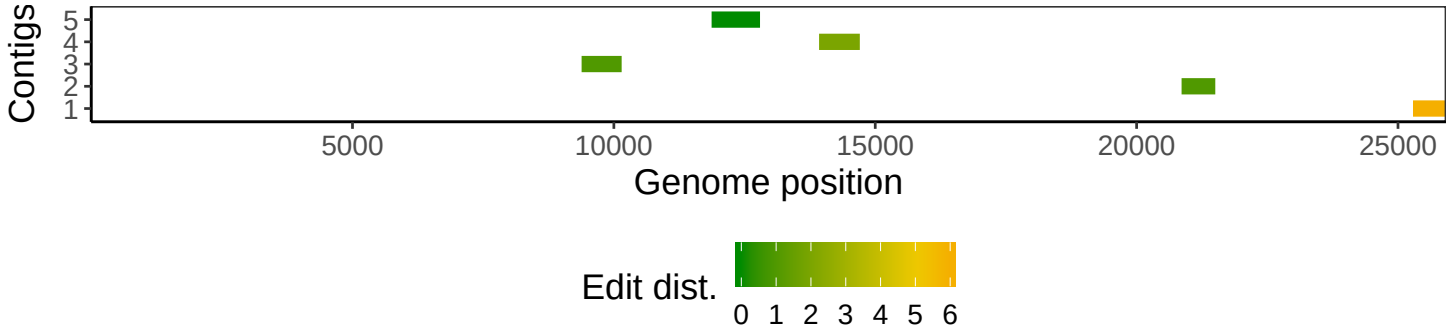
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

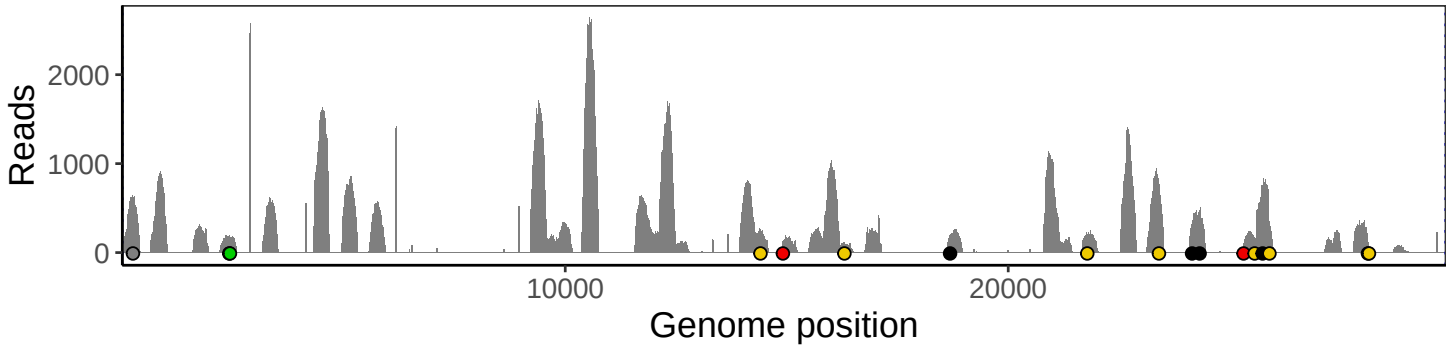


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

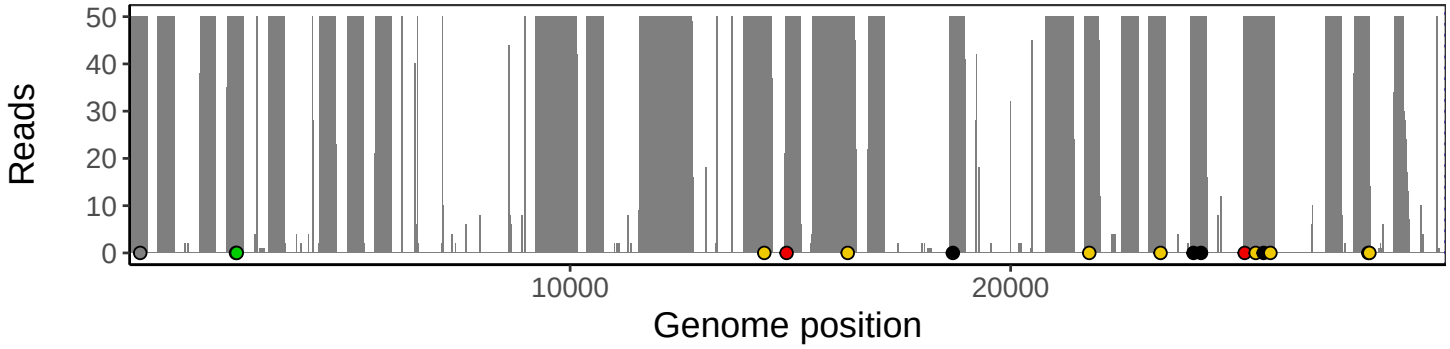


VSP0017-2m | 2020-04-17 | ETA | 219-tri | genomes | single experiment

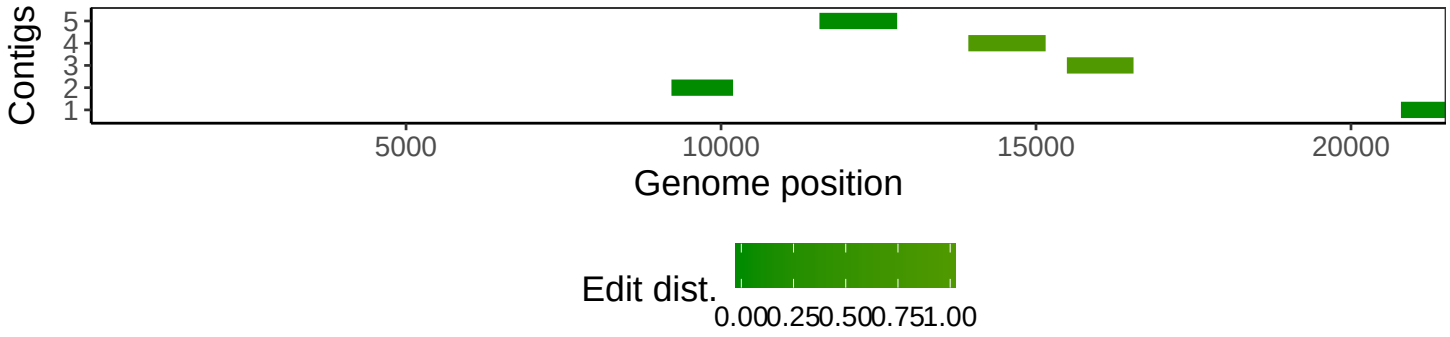
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

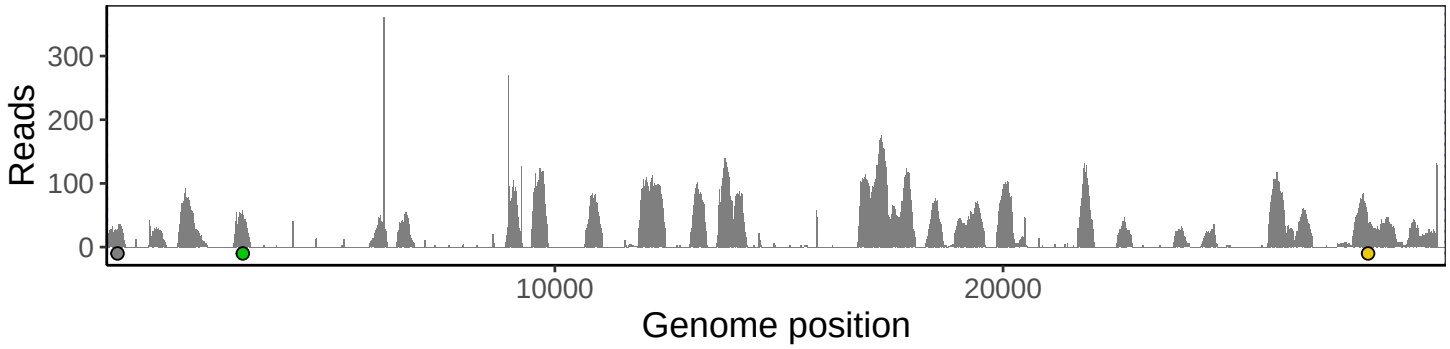


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

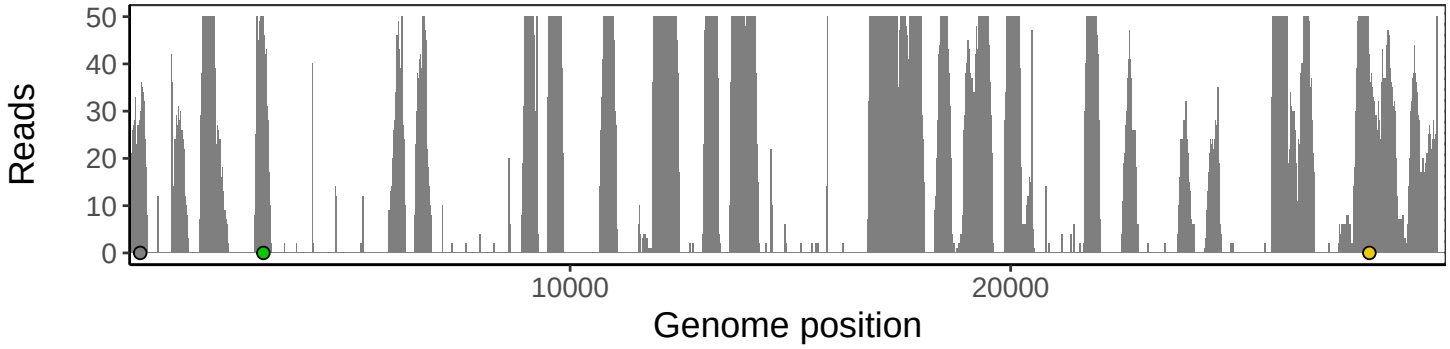


VSP0017-3 | 2020-04-17 | ETA | 219-tri | 9350 genomes | single experiment

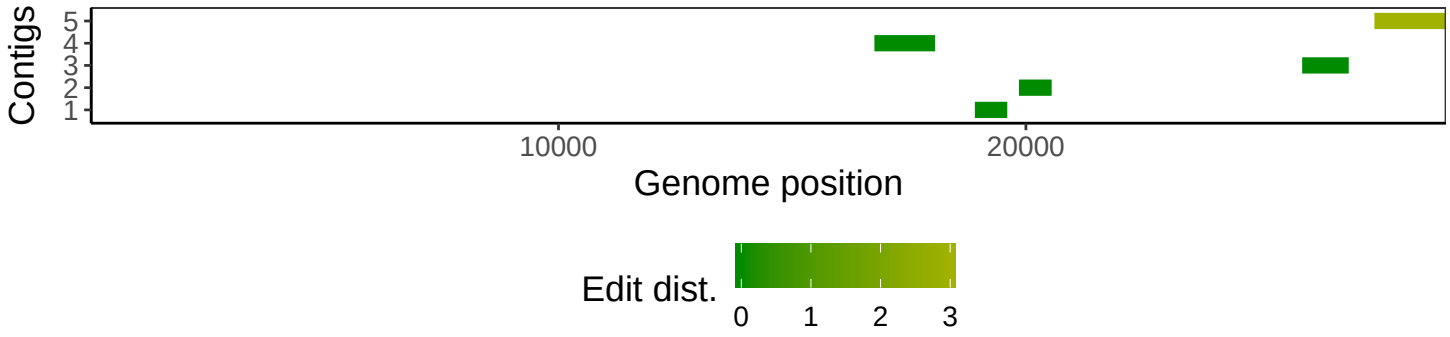
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

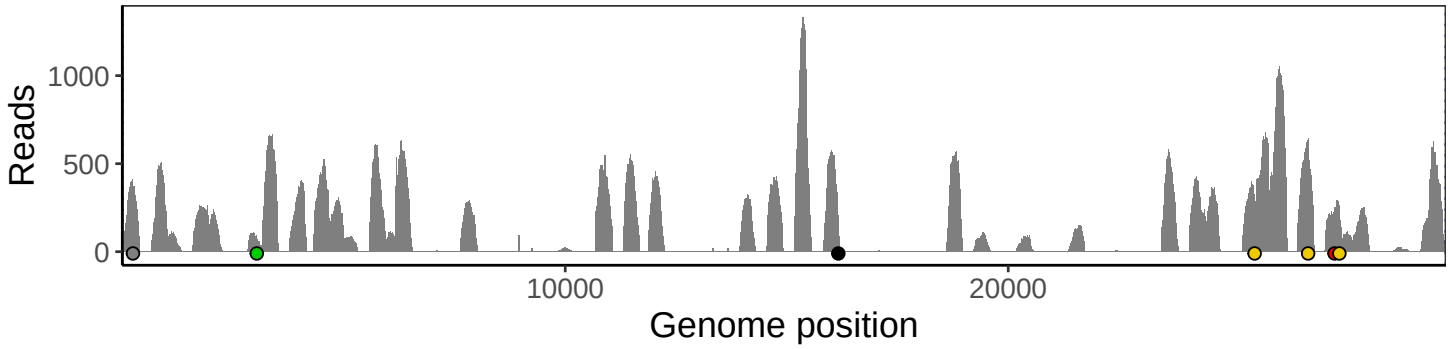


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

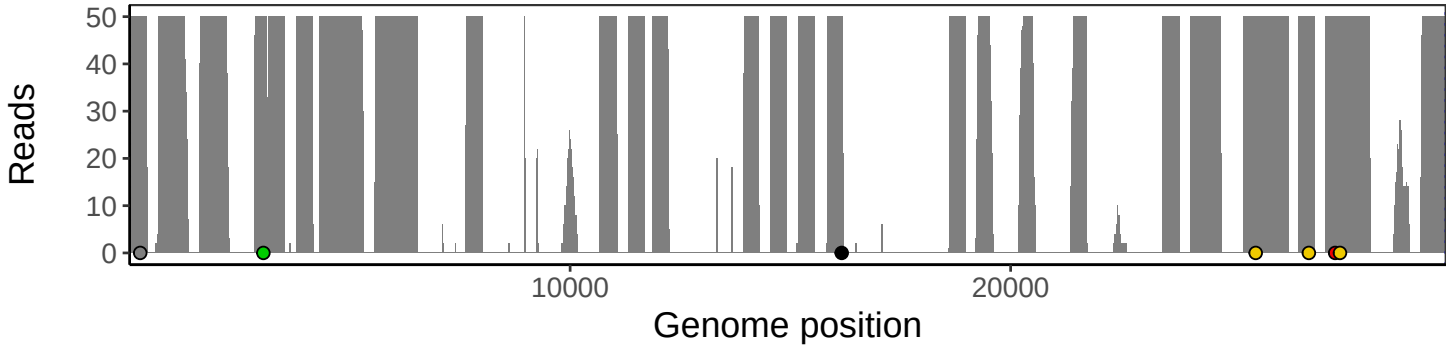


VSP0017-4 | 2020-04-17 | ETA | 219-tri | 1870 genomes | single experiment

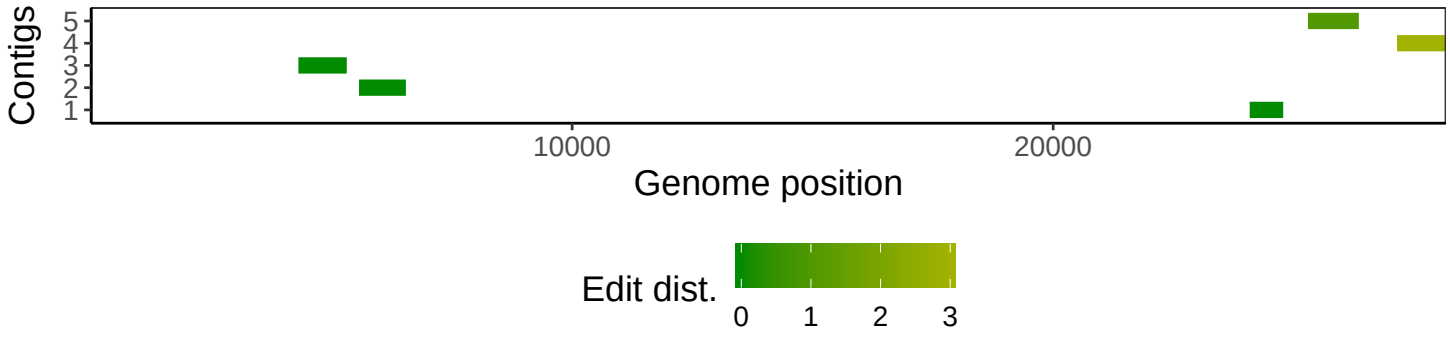
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

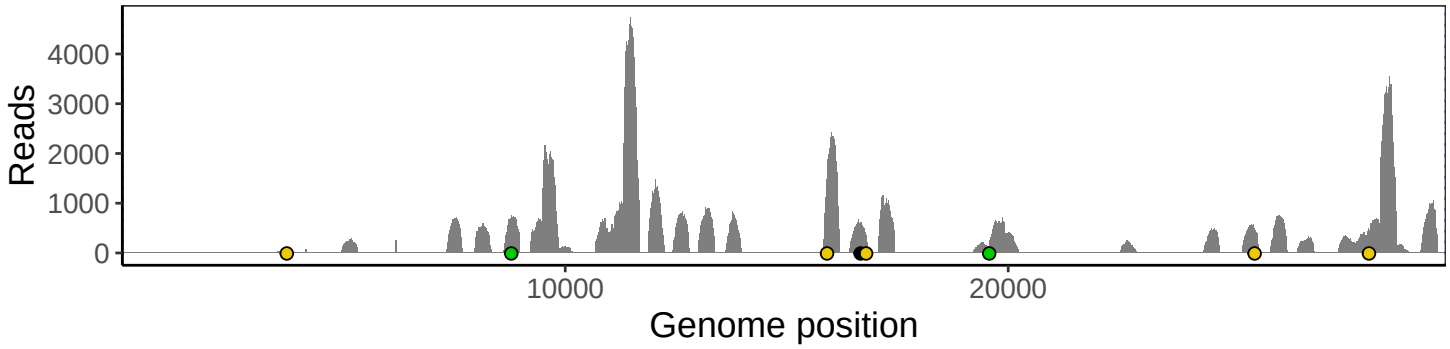


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

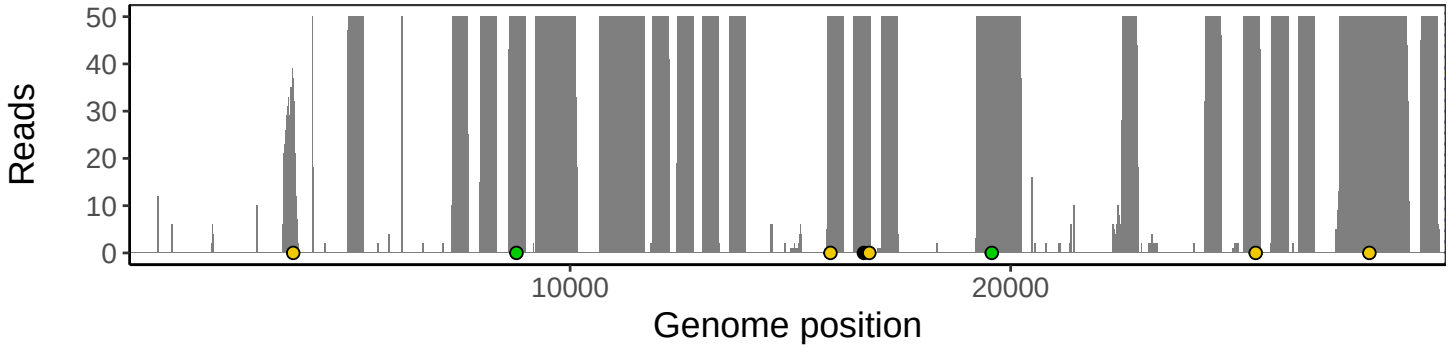


VSP0017-5 | 2020-04-17 | ETA | 219-tri | 1870 genomes | single experiment

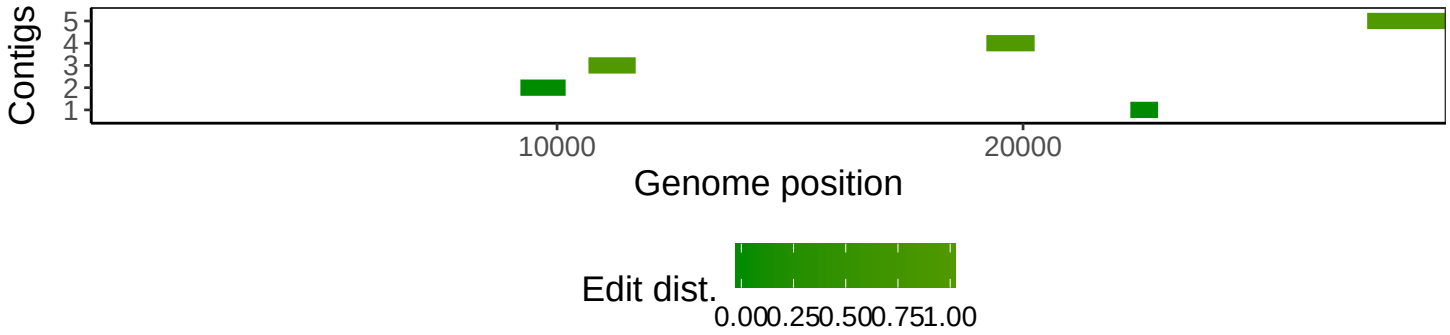
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

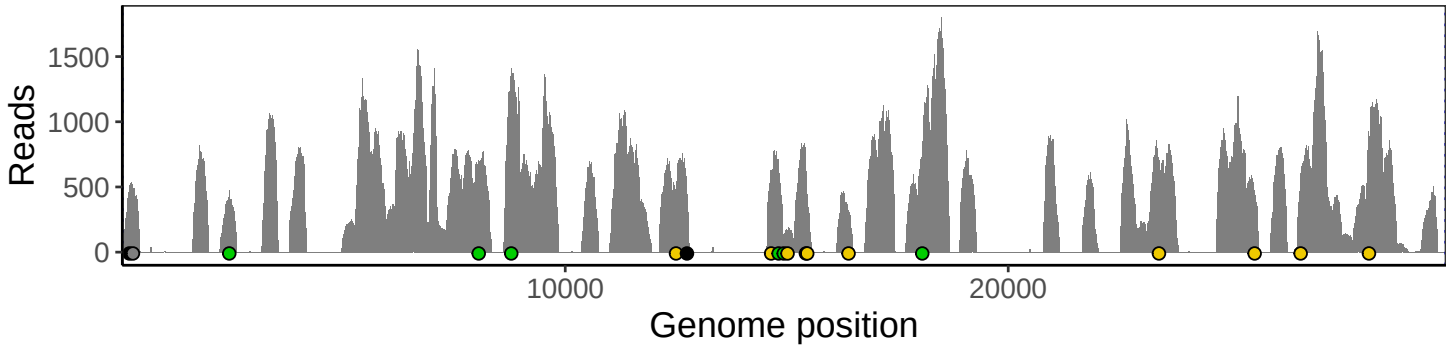


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

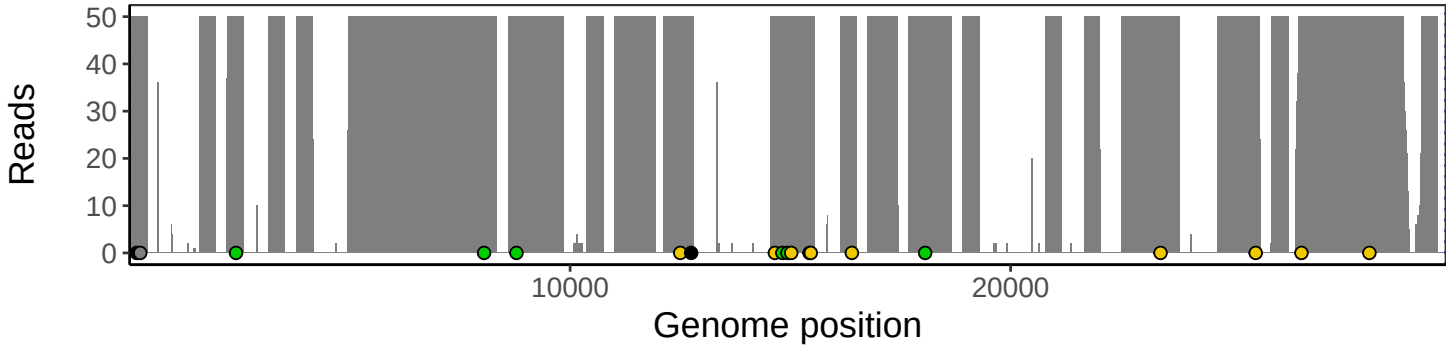


VSP0017-6 | 2020-04-17 | ETA | 219-tri | 1870 genomes | single experiment

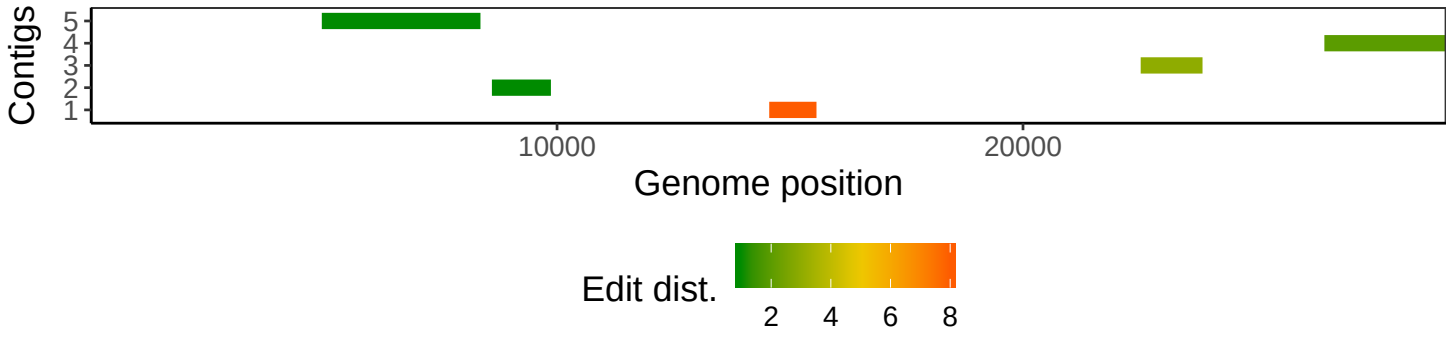
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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