COVID-19 subject AHS21001204

2021-05-05

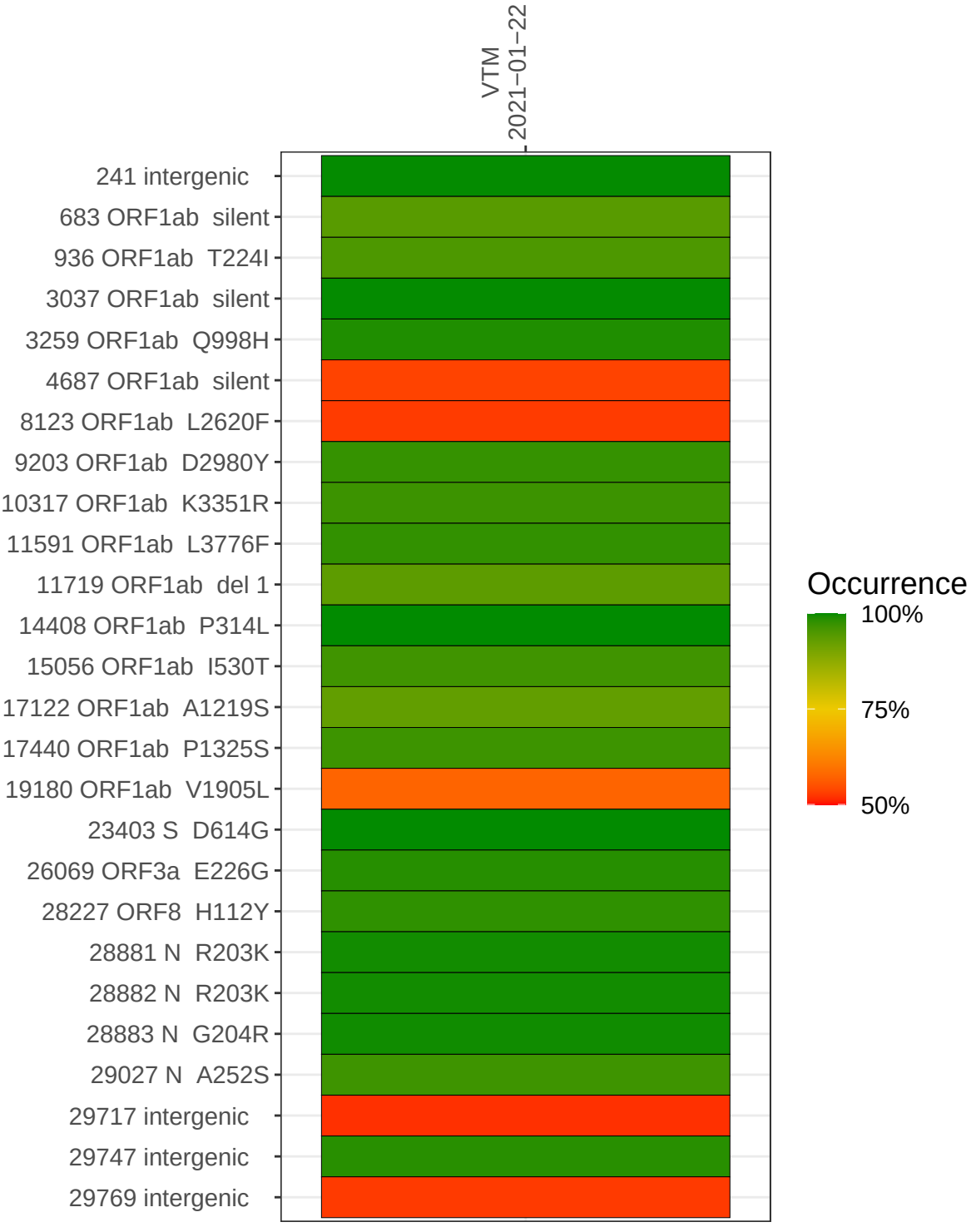
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

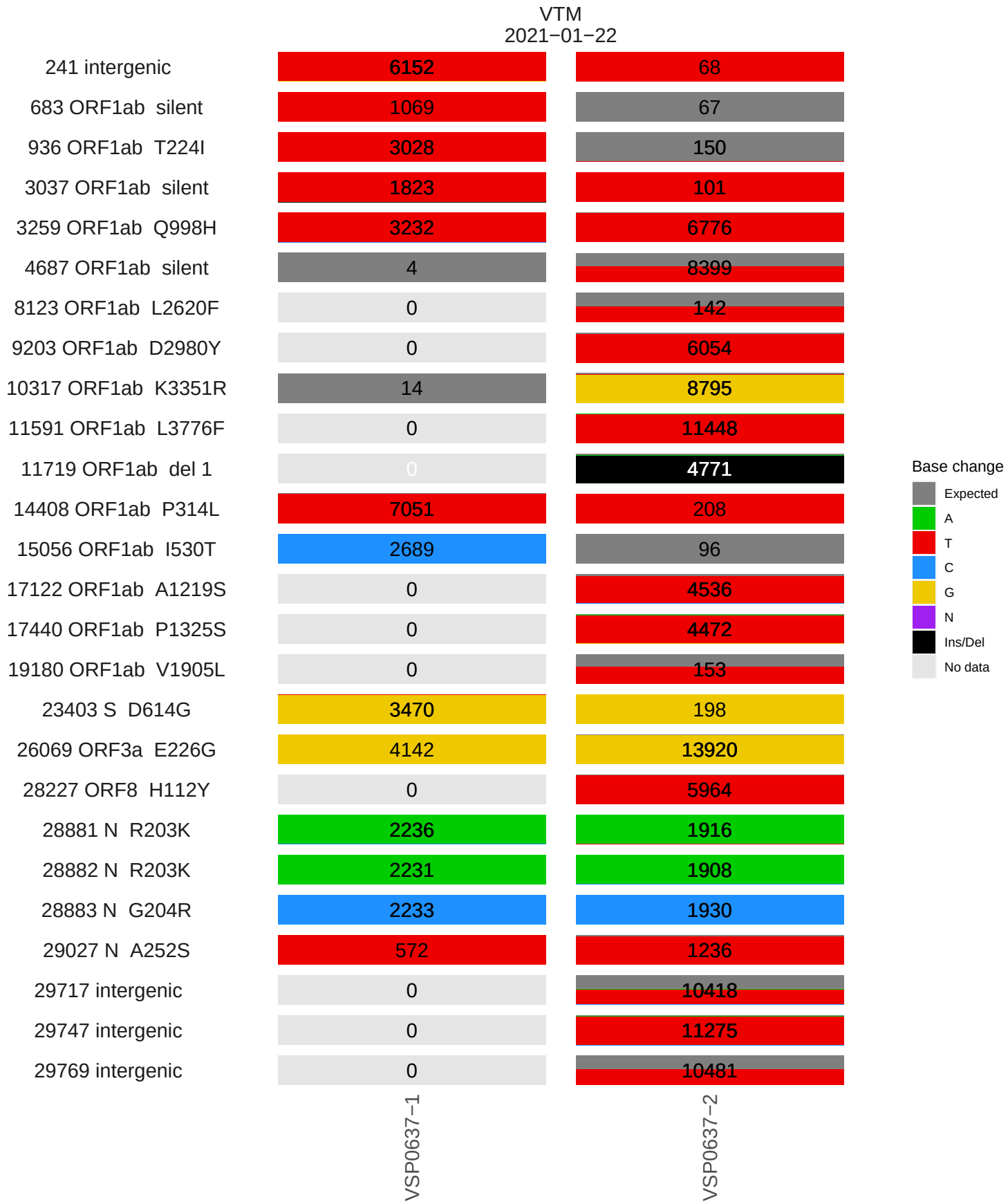
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0637	composite	NA	VTM	2021-01-22	29.93	B.1.1.304	99.9%	99.9%
VSP0637-1	single experiment	NA	VTM	2021-01-22	1.08	NA	36.5%	34.9%
VSP0637-2	single experiment	NA	VTM	2021-01-22	29.82	B.1.1.434	99.8%	99.8%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/Wuhan-Hu-1) 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.

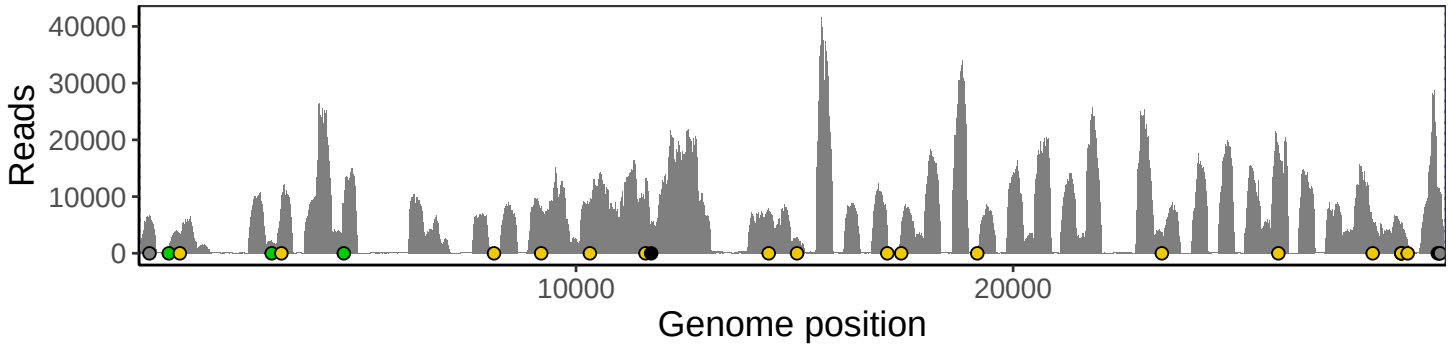



VTM
2021−01−22
241 intergenic
683 ORF1ab silent
936 ORF1ab T224I
3037 ORF1ab silent
3259 ORF1ab Q998H
4687 ORF1ab silent
8123 ORF1ab L2620F
9203 ORF1ab D2980Y
10317 ORF1ab K3351R
11591 ORF1ab L3776F
11719 ORF1ab del 1
14408 ORF1ab P314L
15056 ORF1ab I530T
17122 ORF1ab A1219S
17440 ORF1ab P1325S
19180 ORF1ab V1905L
23403 S D614G
26069 ORF3a E226G
28227 ORF8 H112Y
28881 N R203K
28882 N R203K
28883 N G204R
29027 N A252S
29717 intergenic
29747 intergenic
29769 intergenic
6152
1069
3028
1823
3232
4
0
0
14
0
0
7051
2689
0
0
0
3470
4142
0
2236
2231
2233
572
0
0
0
68
67
150
101
6776
8399
142
6054
8795
11448
4771
208
96
4536
4472
153
198
13920
5964
1916
1908
1930
1236
10418
11275
10481
VSP0637−1
VSP0637−2
Base change
Expected
A
T
C
G
N
Ins/Del
No data

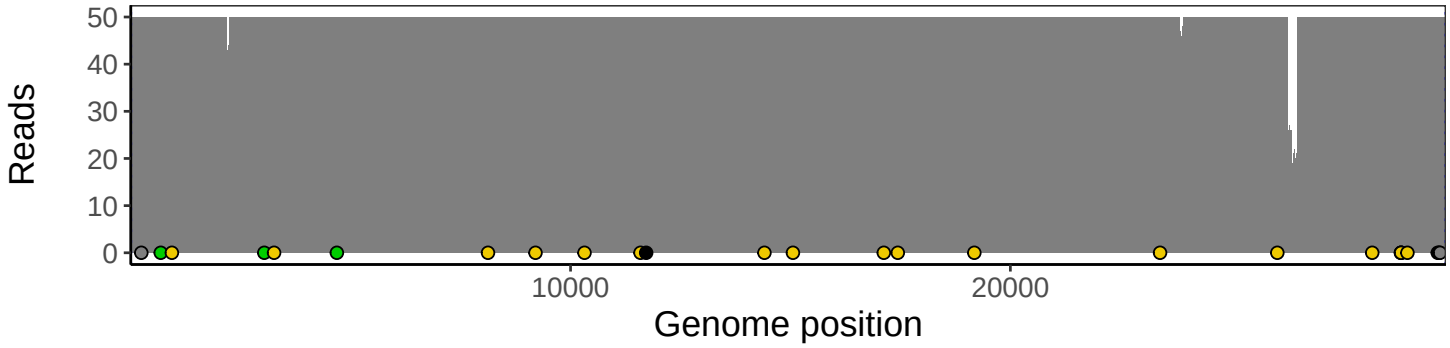
Analyses of individual experiments and composite results

VSP0637 | 2021-01-22 | VTM | H2101200627 | composite result

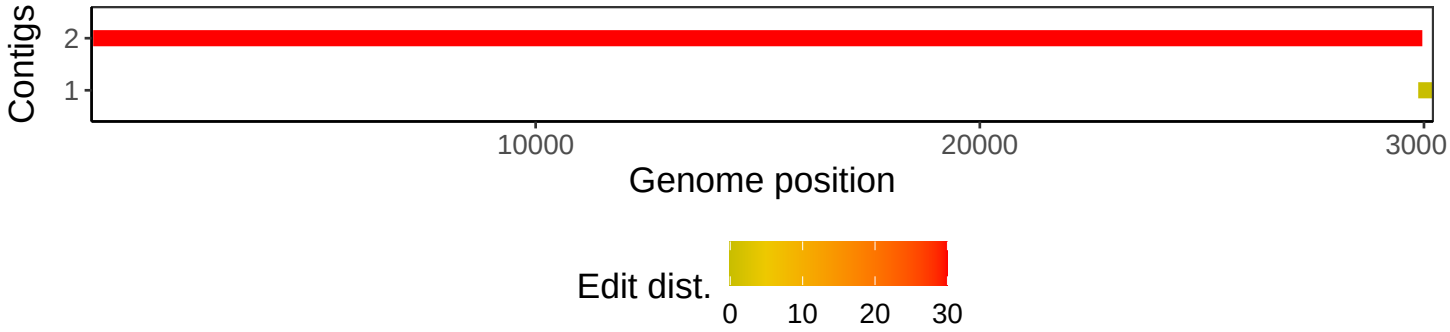
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

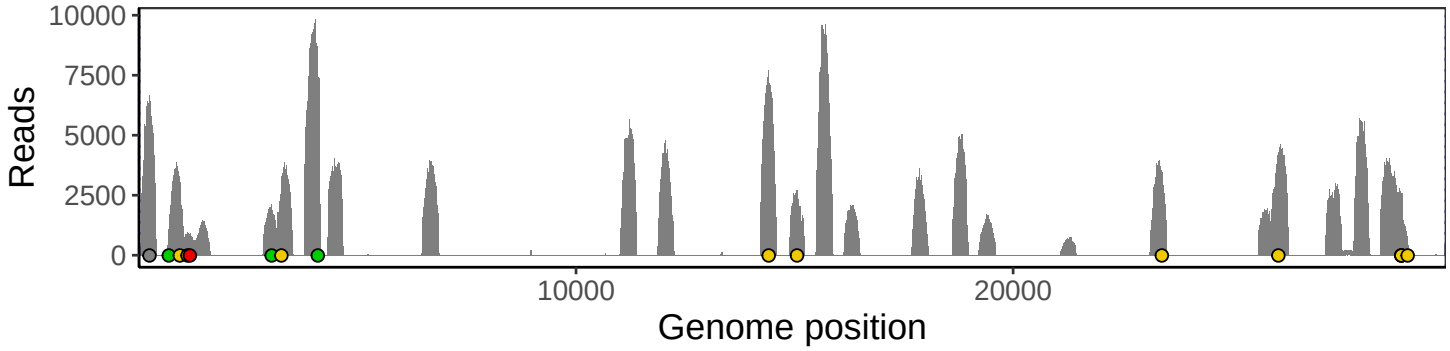


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

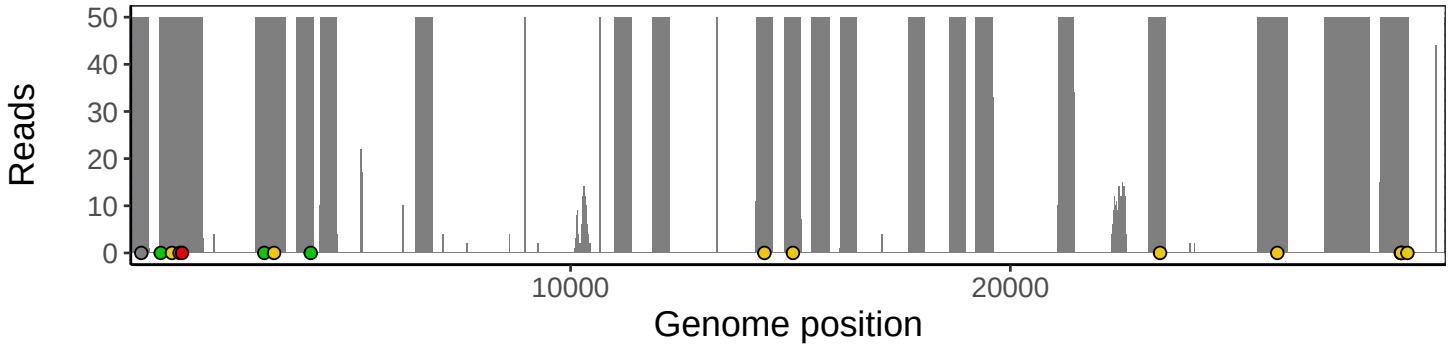


VSP0637-1 | 2021-01-22 | VTM | H2101200627 | genomes | single experiment

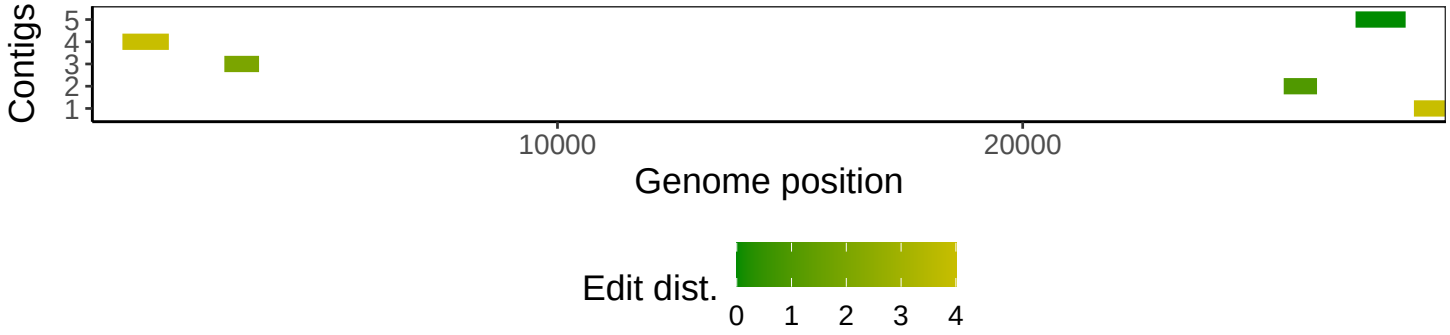
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

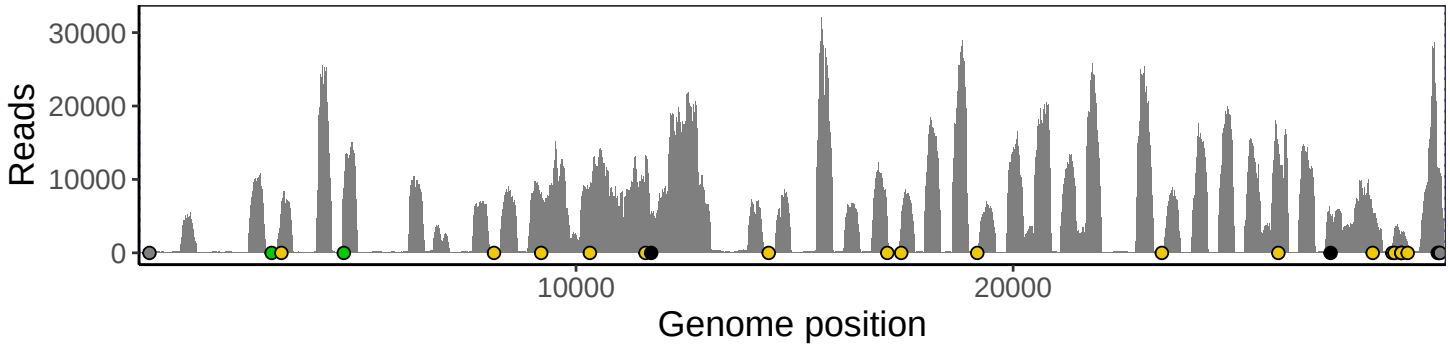


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

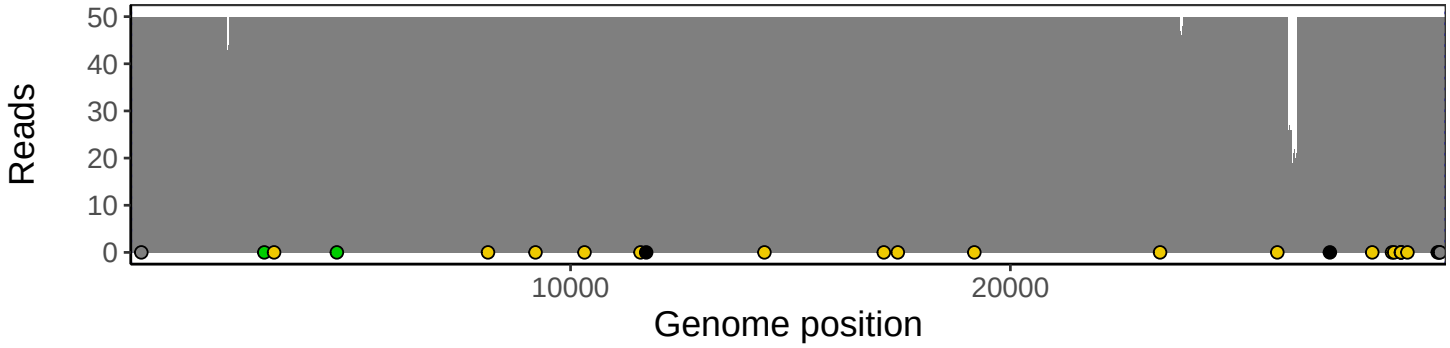


VSP0637-2 | 2021-01-22 | VTM | H2101200627 | genomes | single experiment

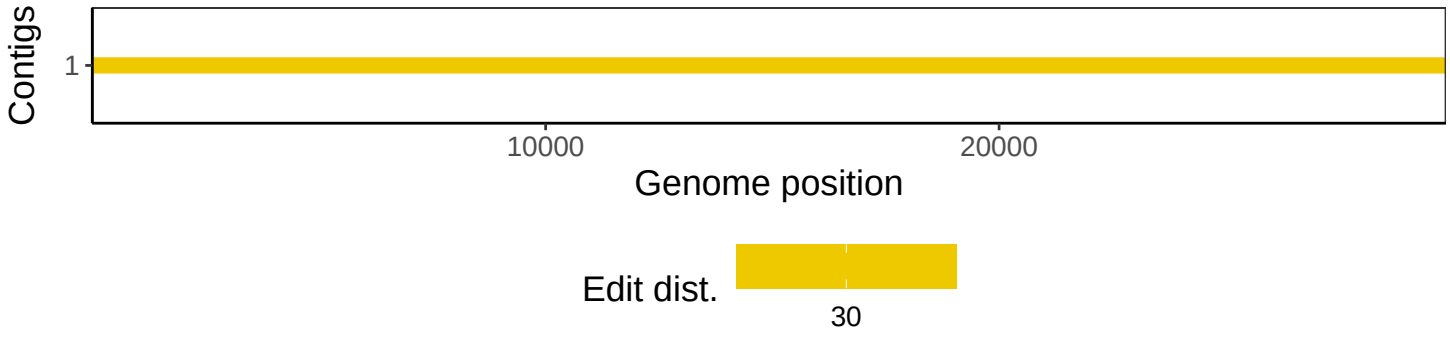
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.3.8
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