COVID-19 subject H2102240764

2021-04-17

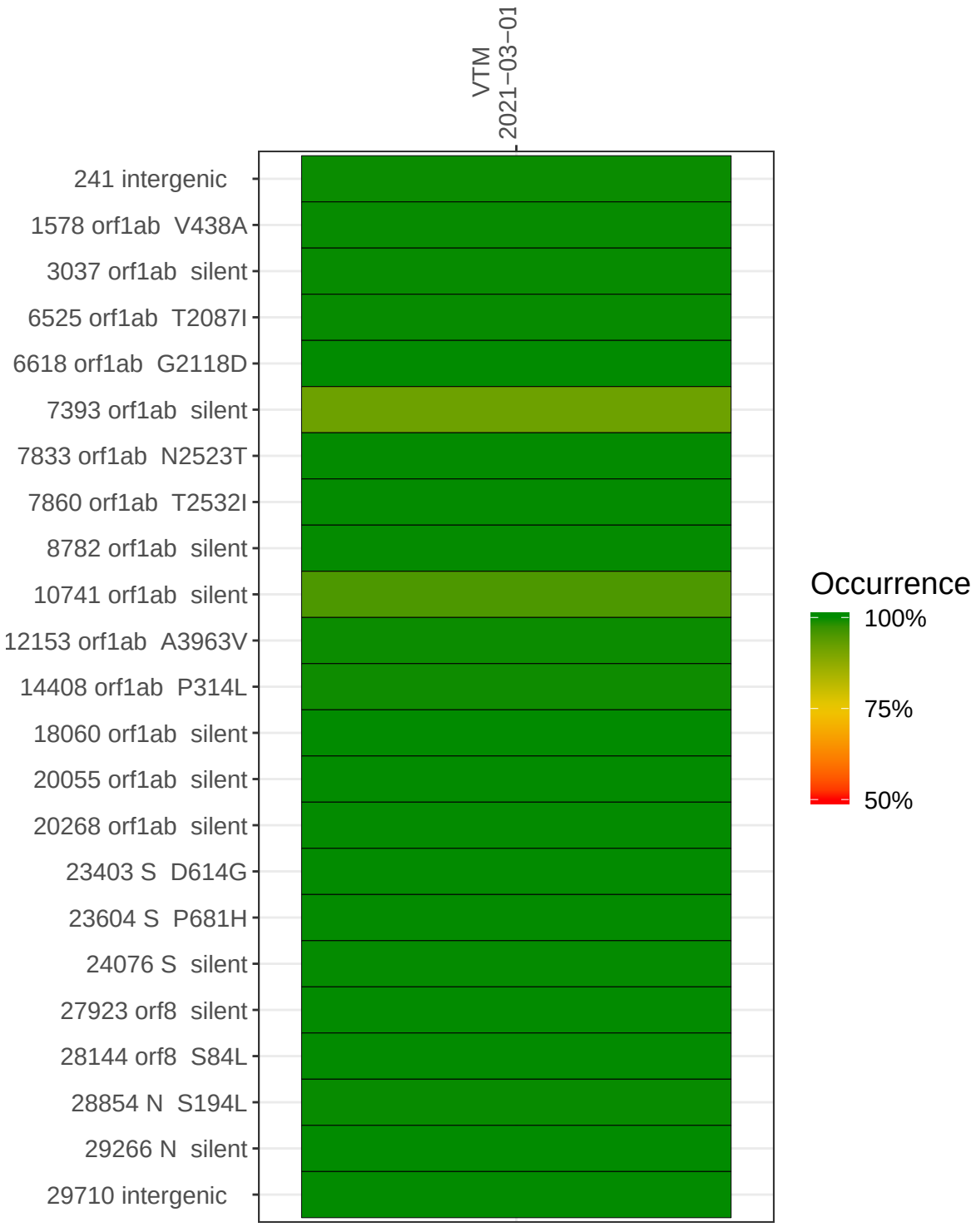
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

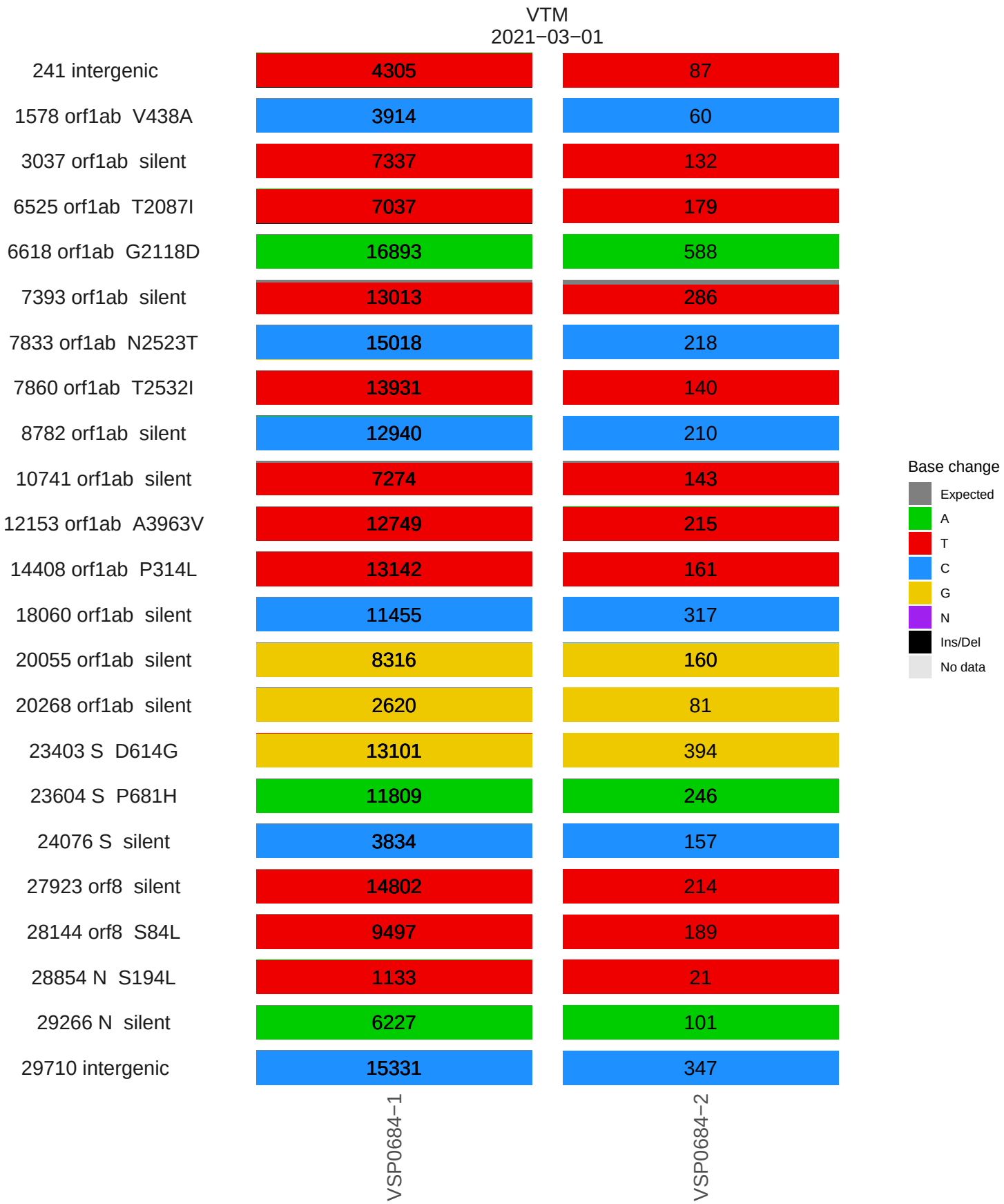
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0684	composite	NA	VTM	2021-03-01	29.94	B.1.243	99.9%	99.9%
VSP0684-1	single experiment	NA	VTM	2021-03-01	29.94	B.1.243	99.9%	99.9%
VSP0684-2	single experiment	NA	VTM	2021-03-01	22.32	B.1.243	99.4%	99.2%

Variants shared across samples

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

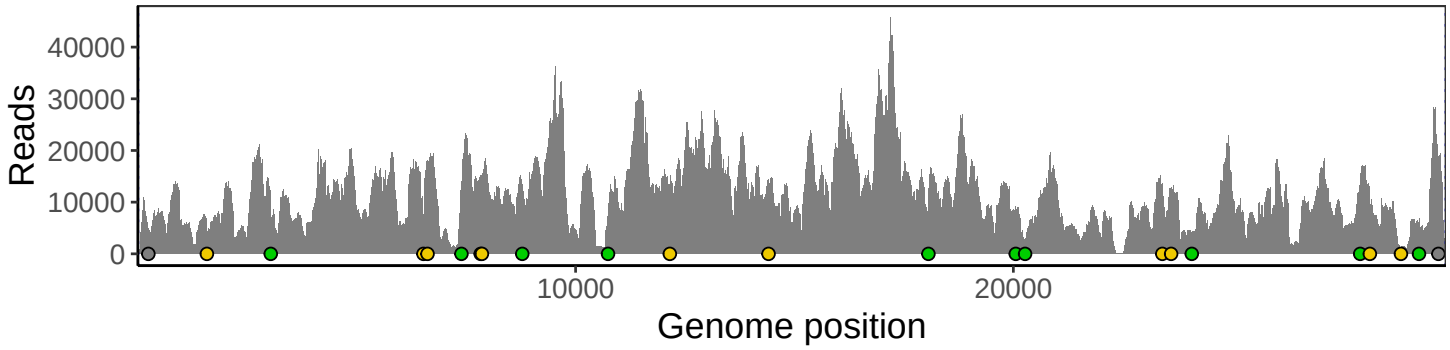



VTM
2021−03−01
241 intergenic
1578 orf1ab V438A
3037 orf1ab silent
6525 orf1ab T2087I
6618 orf1ab G2118D
7393 orf1ab silent
7833 orf1ab N2523T
7860 orf1ab T2532I
8782 orf1ab silent
10741 orf1ab silent
12153 orf1ab A3963V
14408 orf1ab P314L
18060 orf1ab silent
20055 orf1ab silent
20268 orf1ab silent
23403 S D614G
23604 S P681H
24076 S silent
27923 orf8 silent
28144 orf8 S84L
28854 N S194L
29266 N silent
29710 intergenic
4305
3914
7337
7037
16893
13013
15018
13931
12940
7274
12749
13142
11455
8316
2620
13101
11809
3834
14802
9497
1133
6227
15331
87
60
132
179
588
286
218
140
210
143
215
161
317
160
81
394
246
157
214
189
21
101
347
VSP0684−1
VSP0684−2
Base change
Expected
A
T
C
G
N
Ins/Del
No data

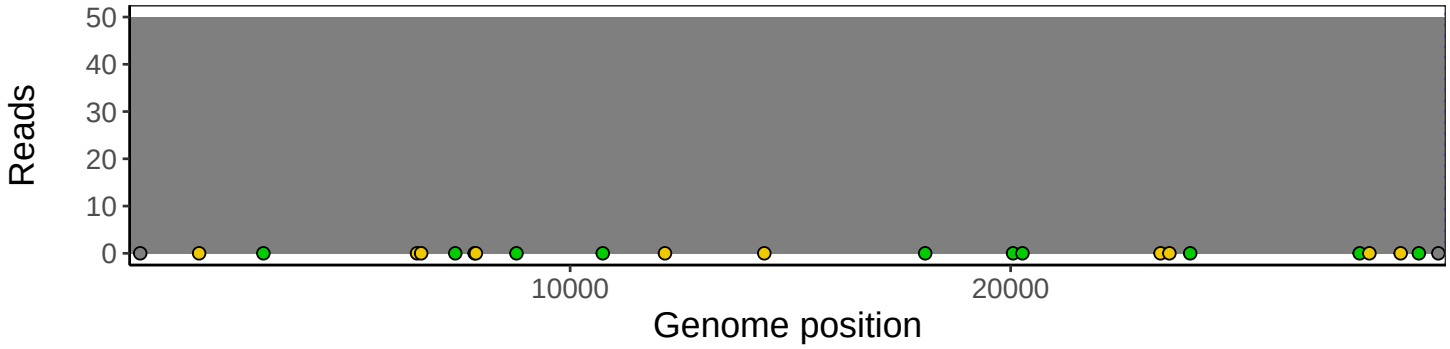
Analyses of individual experiments and composite results

VSP0684 | 2021-03-01 | VTM | H2102240764 | composite result

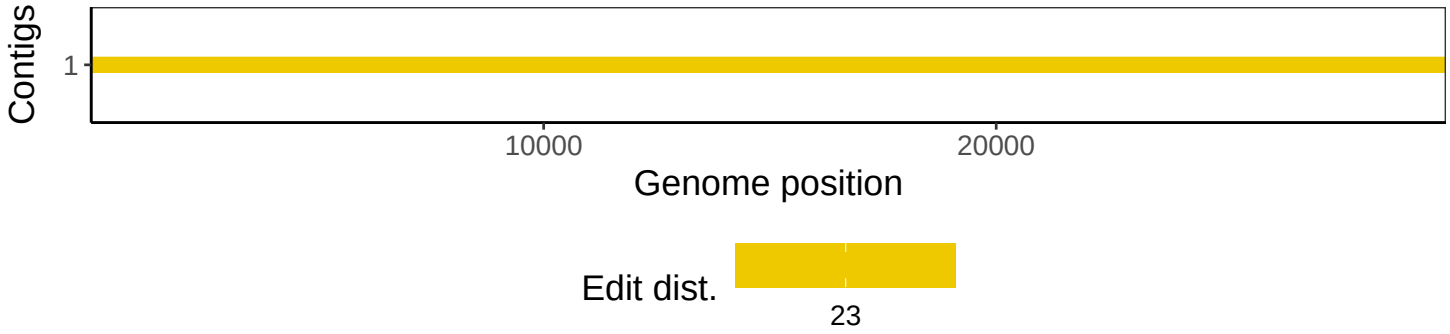
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

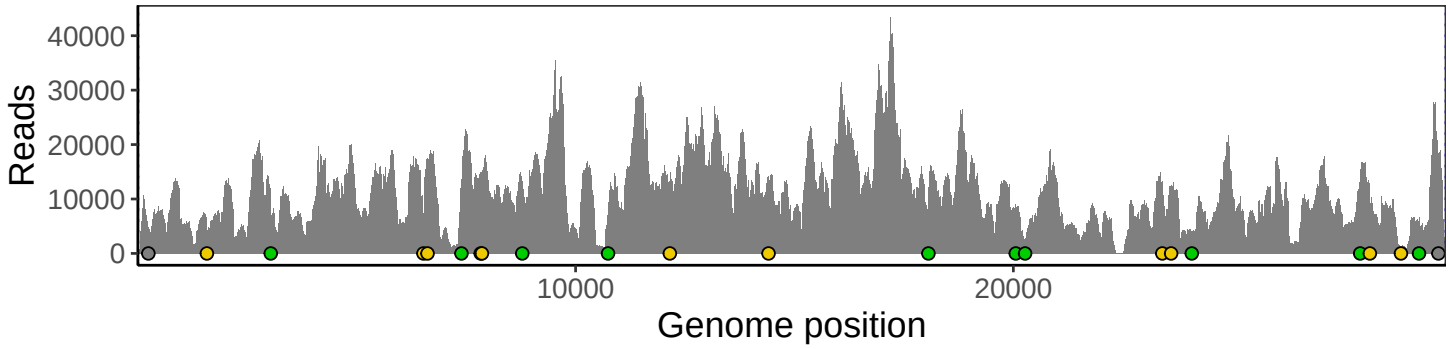


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

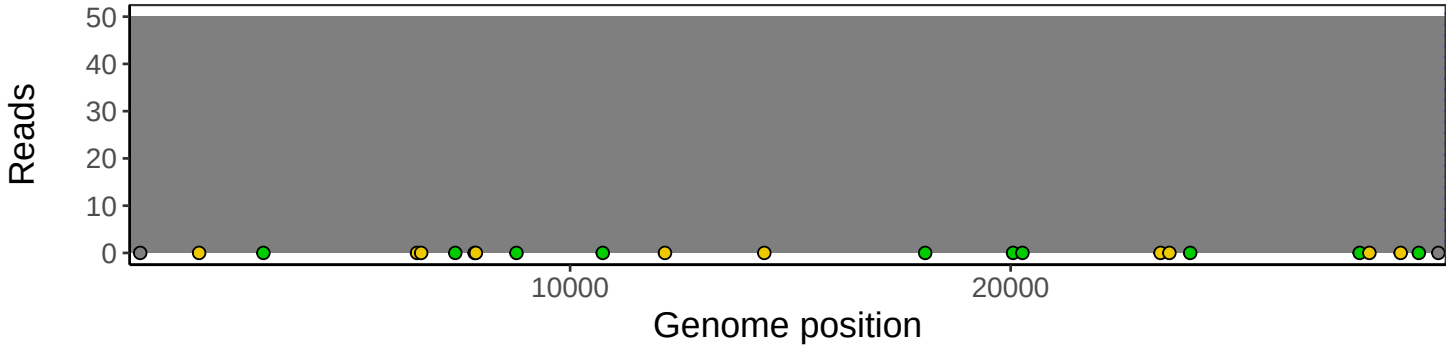


VSP0684-1 | 2021-03-01 | VTM | H2102240764 | genomes | single experiment

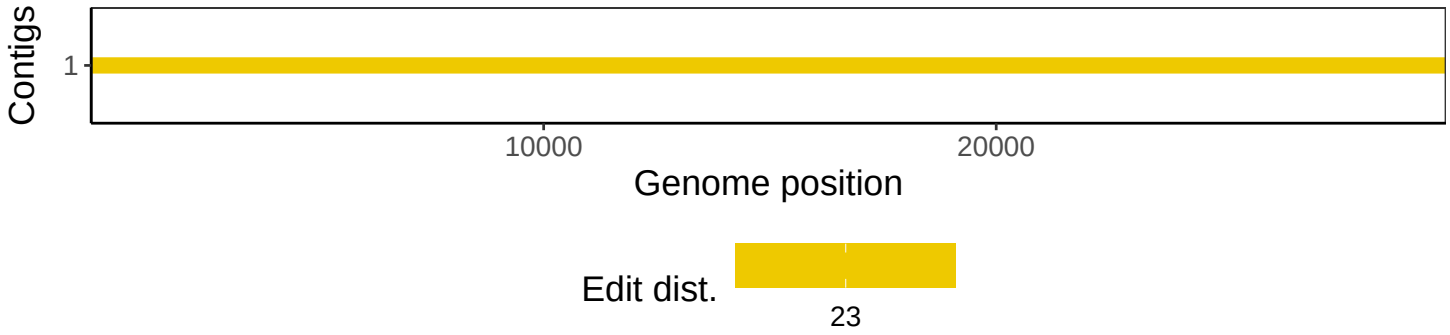
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

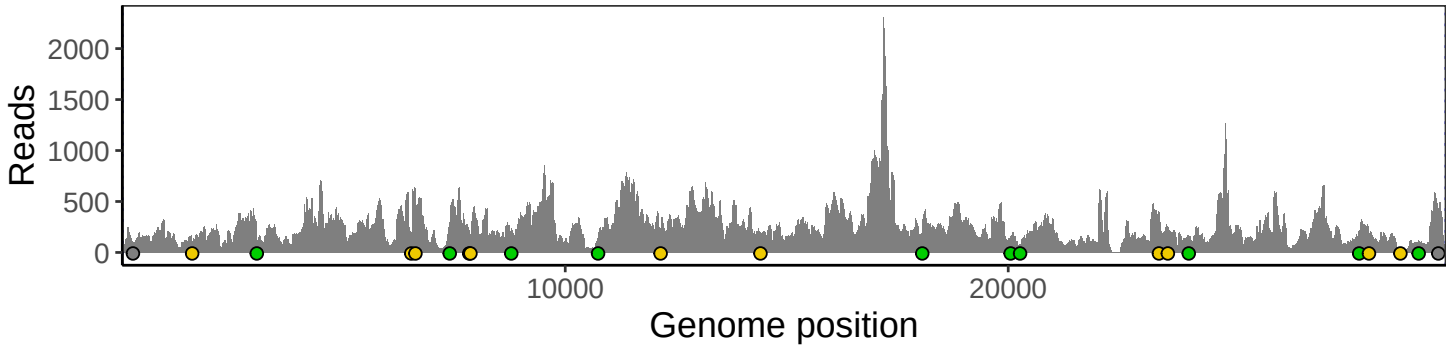


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

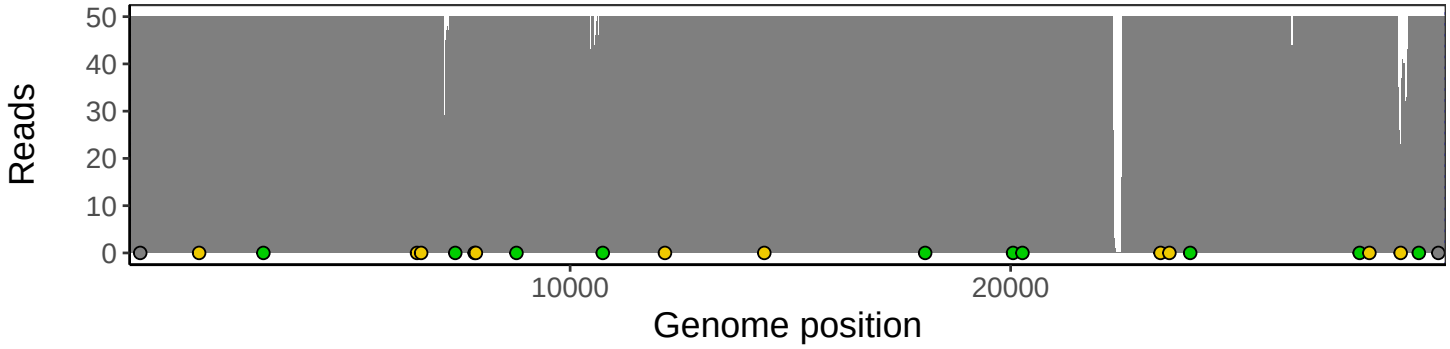


VSP0684-2 | 2021-03-01 | VTM | H2102240764 | genomes | single experiment

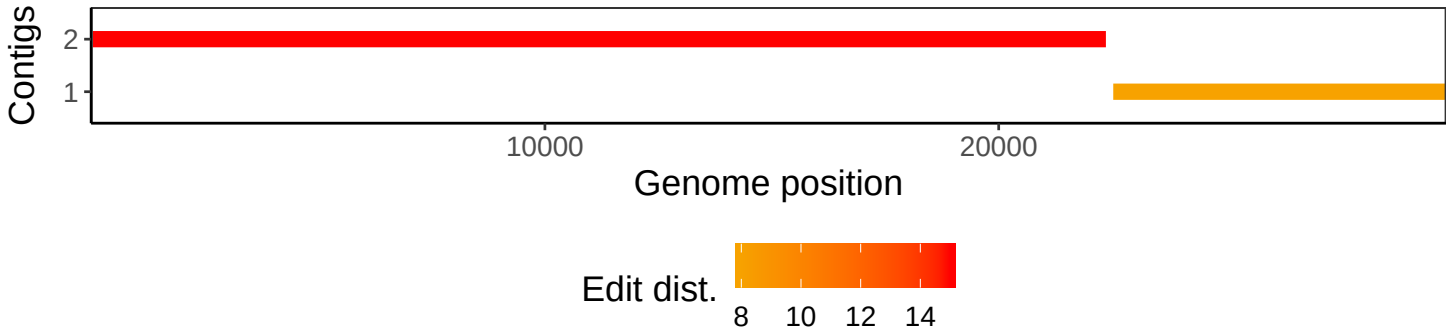
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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