COVID-19 subject 1044999

2021-04-17

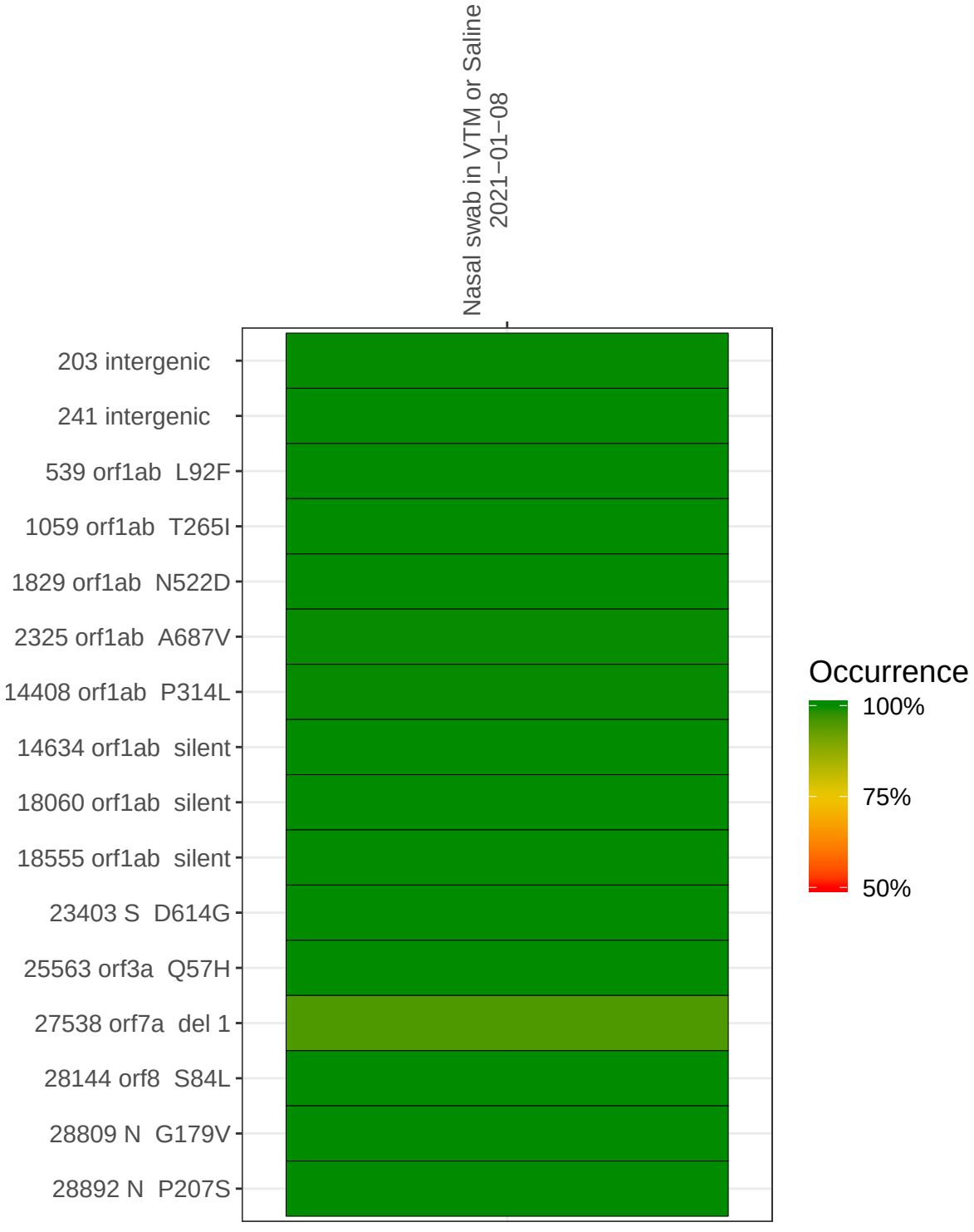
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

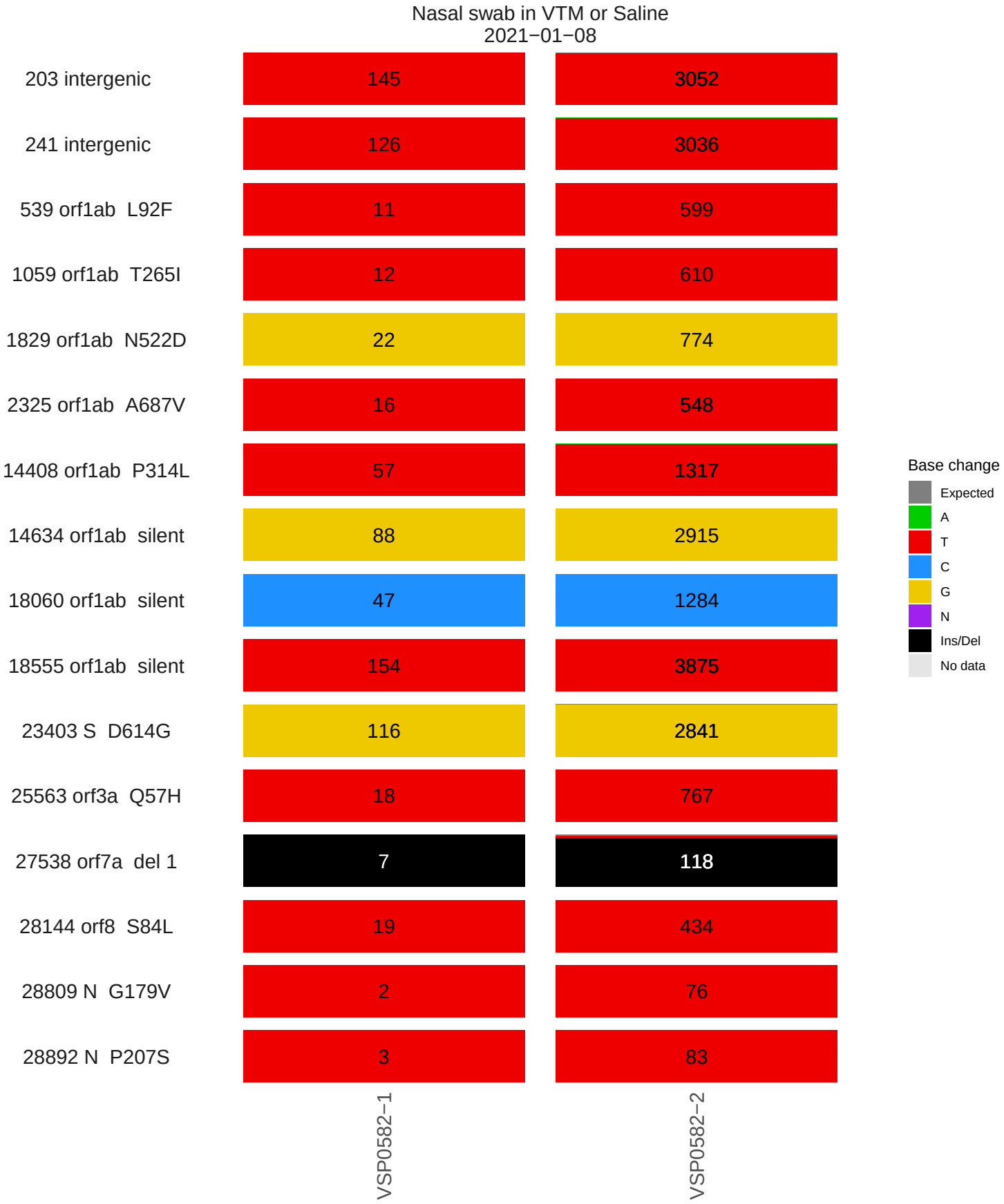
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0582	composite	NA	Nasal swab in VTM or Saline	2021-01-08	4.25	NA	71.7%	69.3%
VSP0582-1	single experiment	NA	Nasal swab in VTM or Saline	2021-01-08	4.24	NA	67.9%	65.8%
VSP0582-2	single experiment	NA	Nasal swab in VTM or Saline	2021-01-08	4.25	NA	71.7%	69.3%

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.

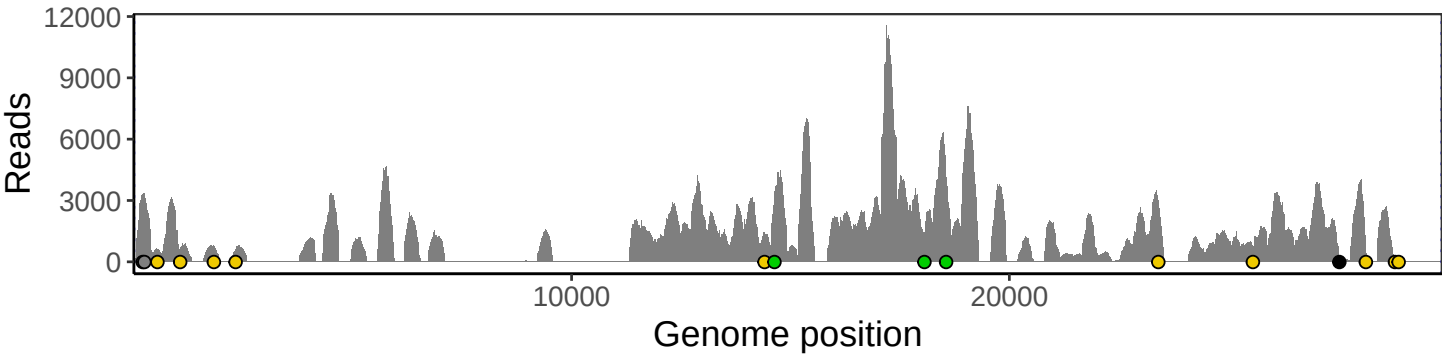



Nasal swab in VTM or Saline
2021−01−08
203 intergenic
241 intergenic
539 orf1ab L92F
1059 orf1ab T265I
1829 orf1ab N522D
2325 orf1ab A687V
14408 orf1ab P314L
14634 orf1ab silent
18060 orf1ab silent
18555 orf1ab silent
23403 S D614G
25563 orf3a Q57H
27538 orf7a del 1
28144 orf8 S84L
28809 N G179V
28892 N P207S
145
126
11
12
22
16
57
88
47
154
116
18
7
19
2
3
3052
3036
599
610
774
548
1317
2915
1284
3875
2841
767
118
434
76
83
VSP0582−1
VSP0582−2
Base change
Expected
A
T
C
G
N
Ins/Del
No data

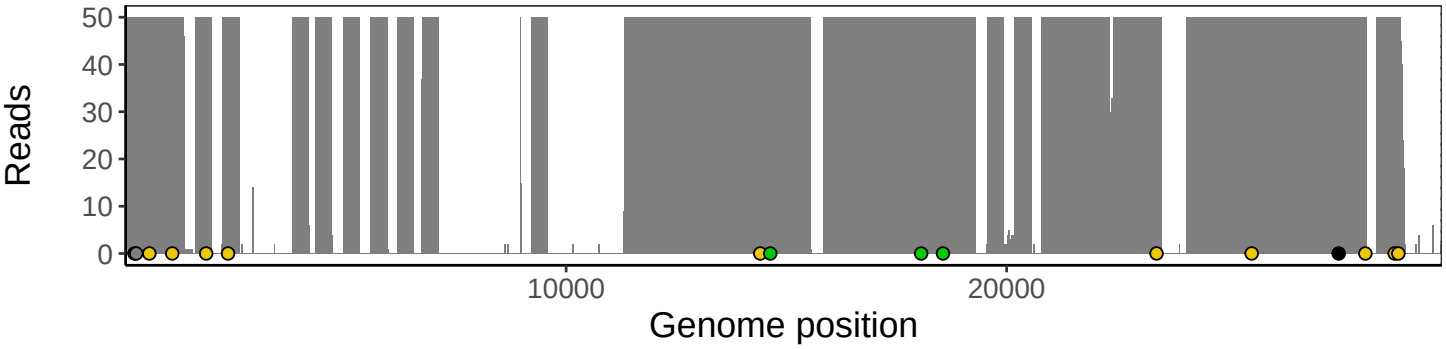
Analyses of individual experiments and composite results

VSP0582 | 2021-01-08 | Nasal swab in VTM or Saline | E50701962-NR; 695215333 | composite result

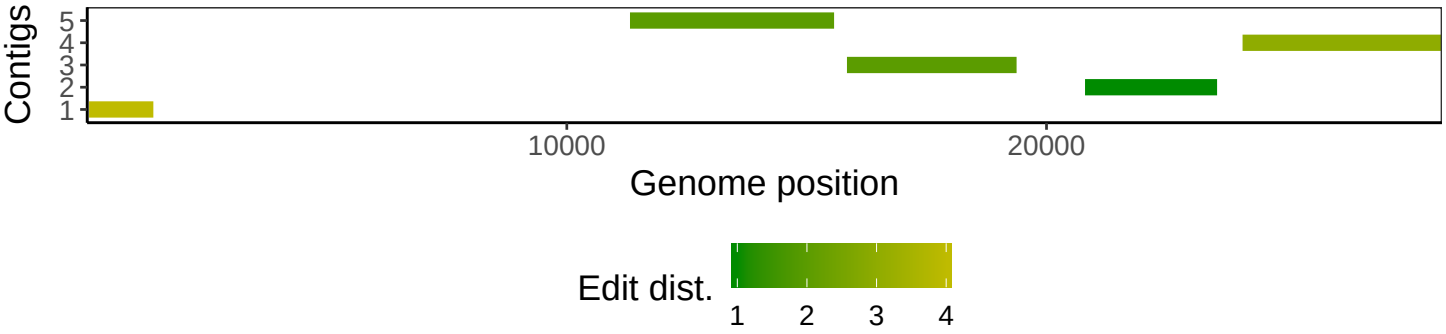
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

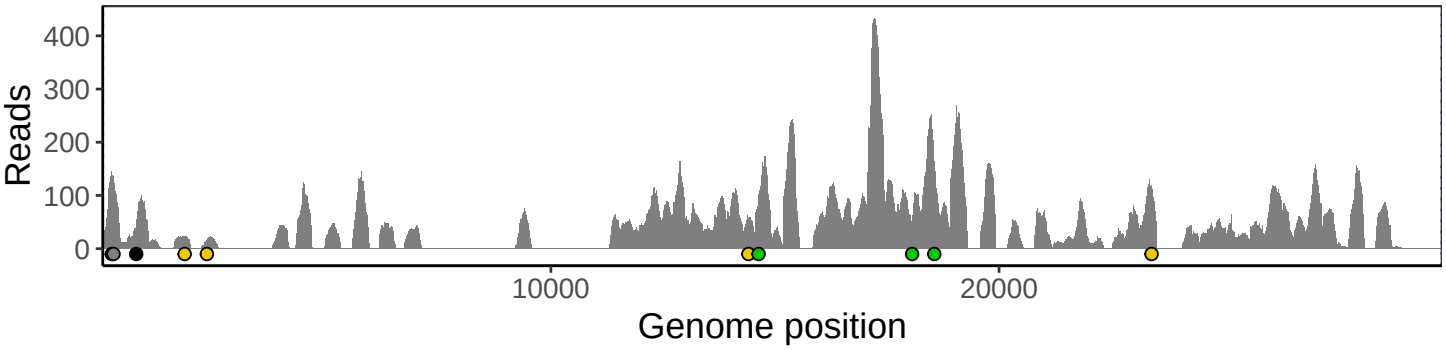


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

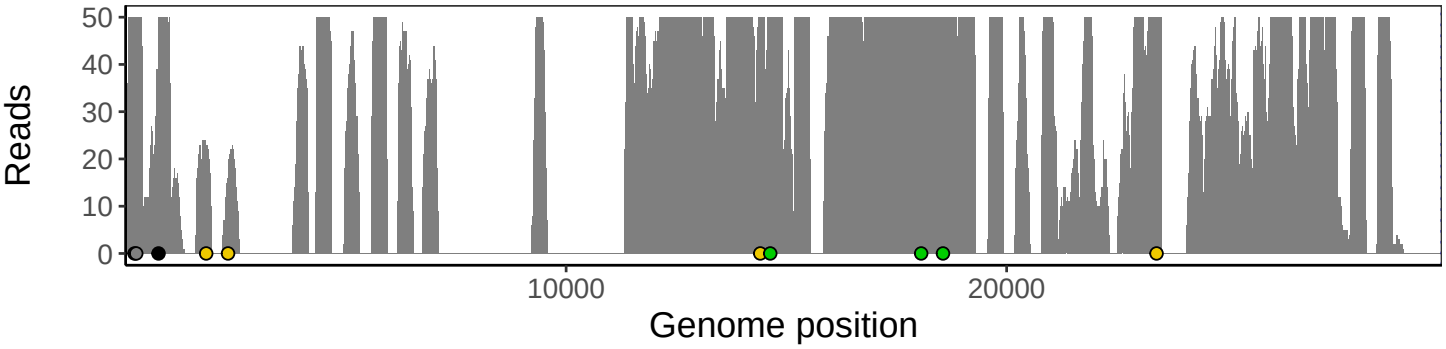


VSP0582-1 | 2021-01-08 | Nasal swab in VTM or Saline | E50701962-NR; 695215333 | genomes | single experiment

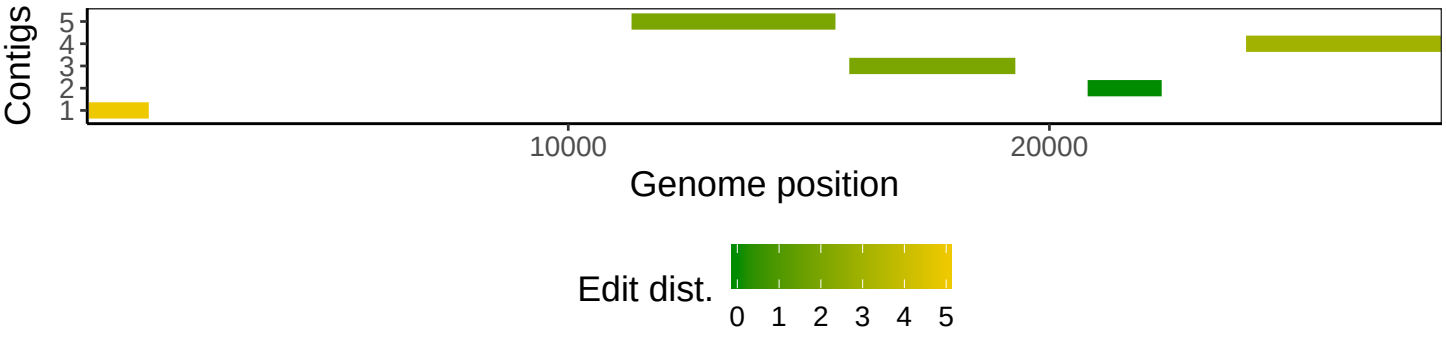
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

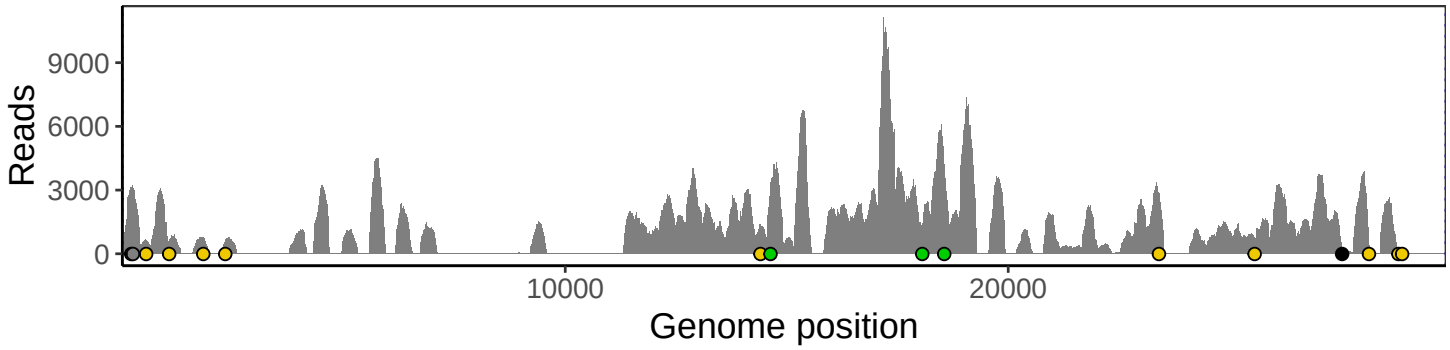


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

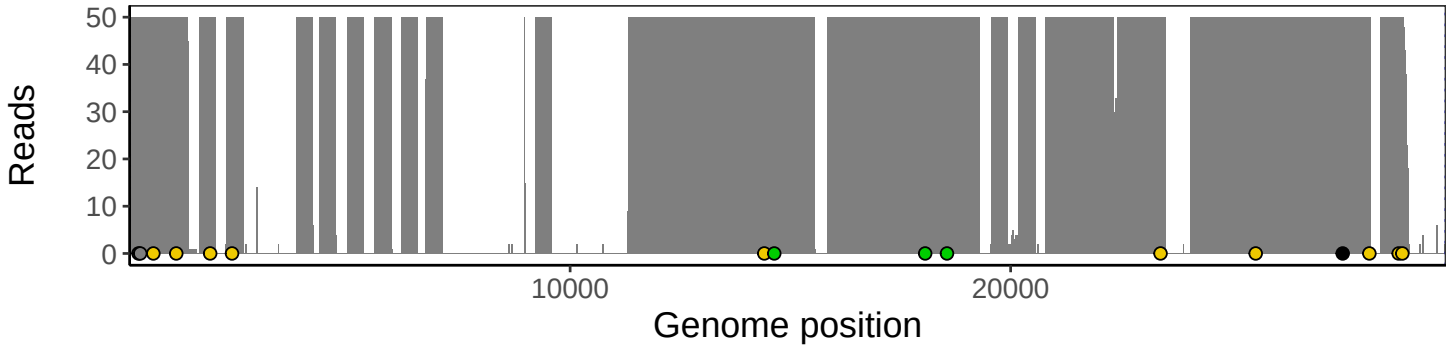


VSP0582-2 | 2021-01-08 | Nasal swab in VTM or Saline | E50701962-NR; 695215333 | genomes | single experiment

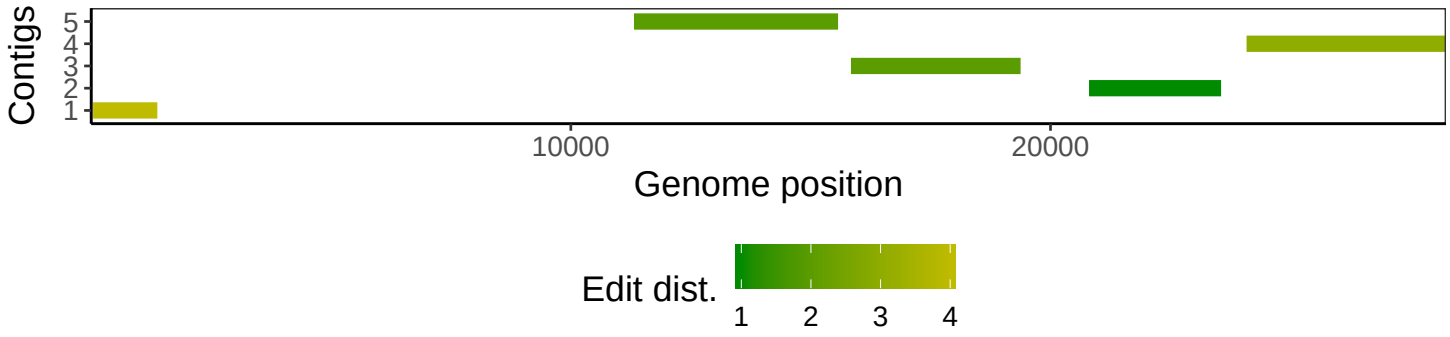
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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