# COVID-19 subject PQ-Seq9

*2021-06-23*

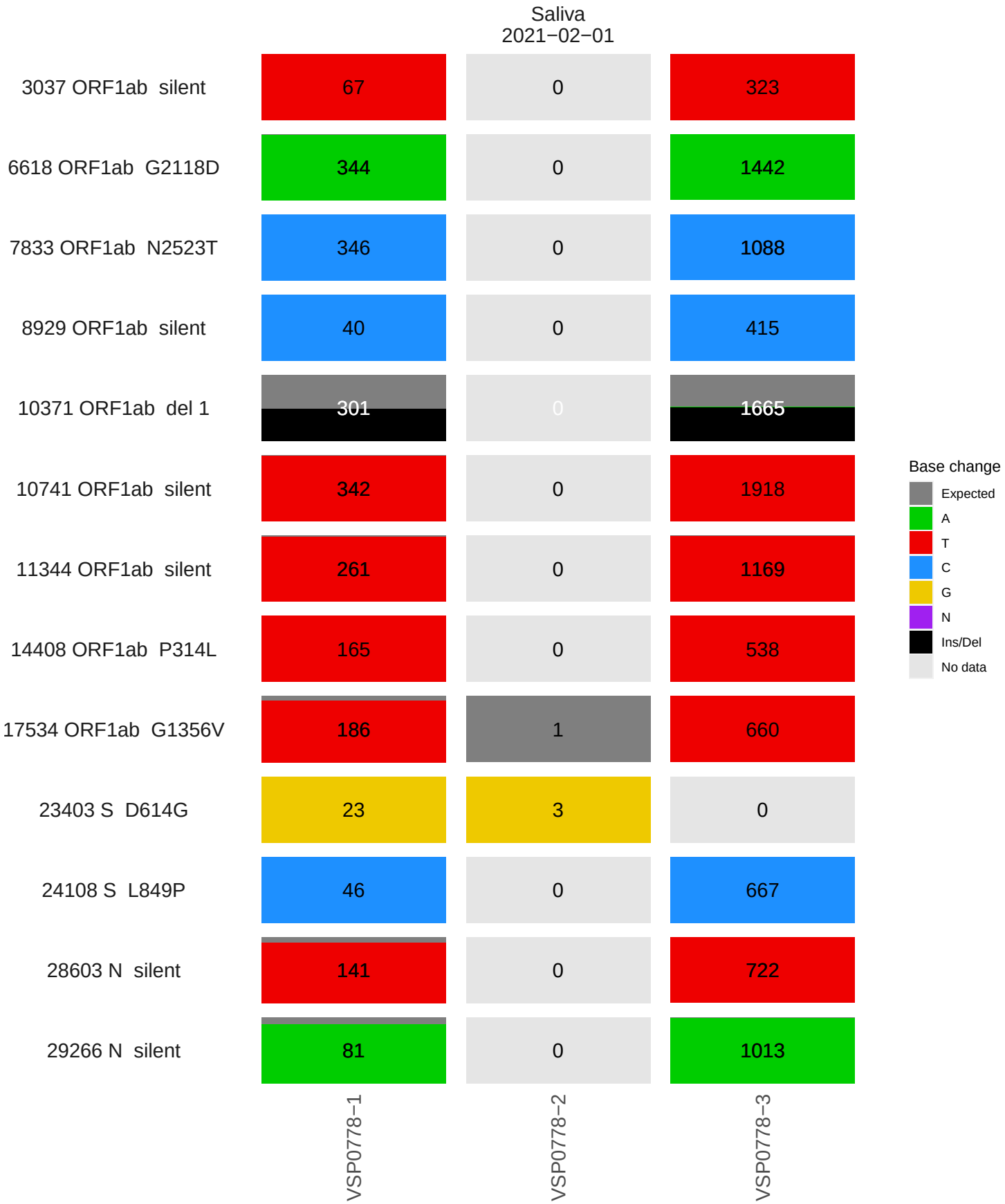
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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) |
|---|---|---|---|---|---|---|---|---|
| VSP0778 | composite | NA | Saliva | 2021-02-01 | 1.21 | NA | 82.2% | 61.4% |
| VSP0778-1 | single experiment | NA | Saliva | 2021-02-01 | 1.02 | NA | 76.2% | 54.1% |
| VSP0778-2 | single experiment | NA | Saliva | 2021-02-01 | 0.37 | NA | 11.7% | 1.6% |
| VSP0778-3 | single experiment | NA | Saliva | 2021-02-01 | 0.60 | NA | 54.0% | 39.6% |

**Variants shared across samples**

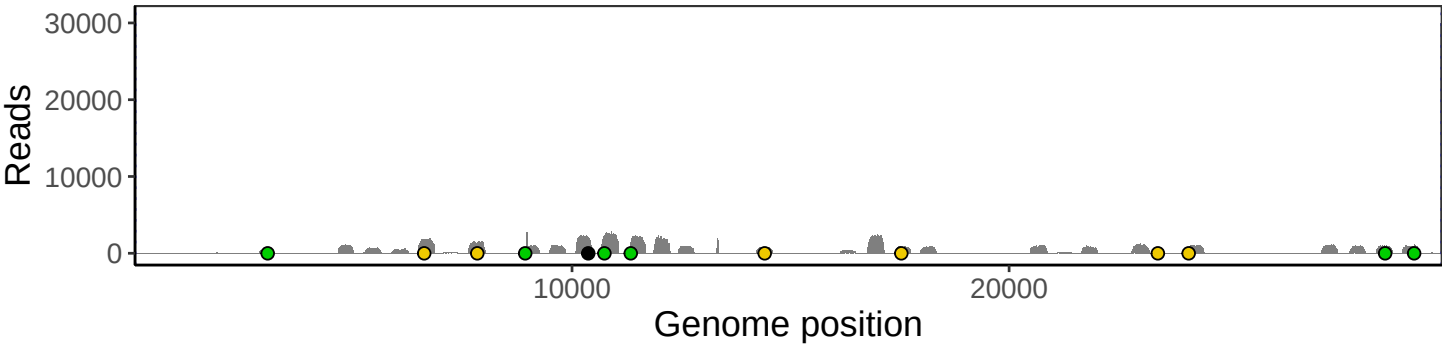
The heat map below shows how variants (reference genome /home/common/SARS-CoV-2-Philadelphia/Wuhan-Hu-1) are shared across subject samples where the percent variance is colored. Variants are called if a variant position is covered by 5 or more reads, the alternative base is found in $> 50\%$ of read 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.


Saliva
2021−02−01
3037 ORF1ab silent
6618 ORF1ab G2118D
7833 ORF1ab N2523T
8929 ORF1ab silent
10371 ORF1ab del 1
10741 ORF1ab silent
11344 ORF1ab silent
14408 ORF1ab P314L
17534 ORF1ab G1356V
23403 S D614G
24108 S L849P
28603 N silent
29266 N silent
67
0
323
344
0
1442
346
0
1088
40
0
415
301
0
1665
342
0
1918
261
0
1169
165
0
538
186
1
660
23
3
0
46
0
667
141
0
722
81
0
1013
VSP0778−1
VSP0778−2
VSP0778−3
Base change
Expected
A
T
C
G
N
Ins/Del
No data

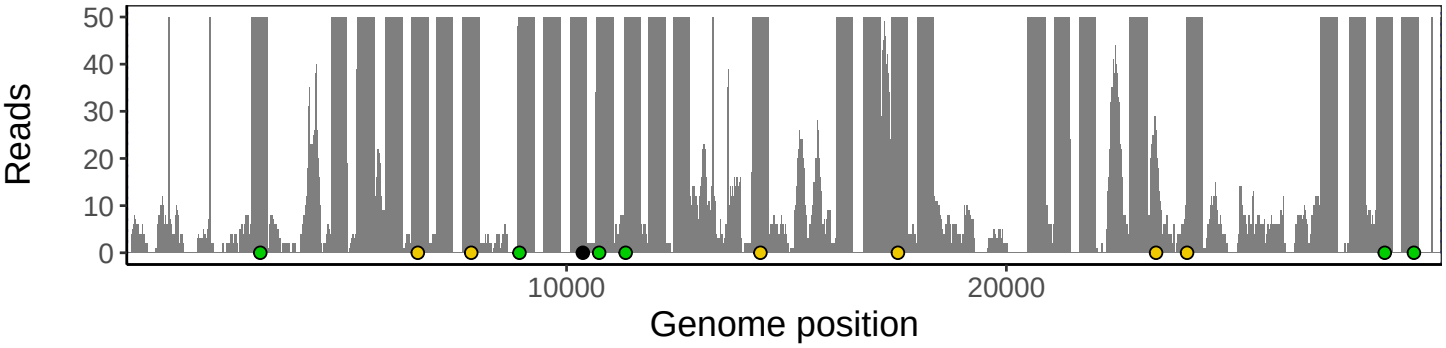
## Analyses of individual experiments and composite results

### VSP0778 | 2021-02-01 | Saliva | PQ-Seq9 | 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.



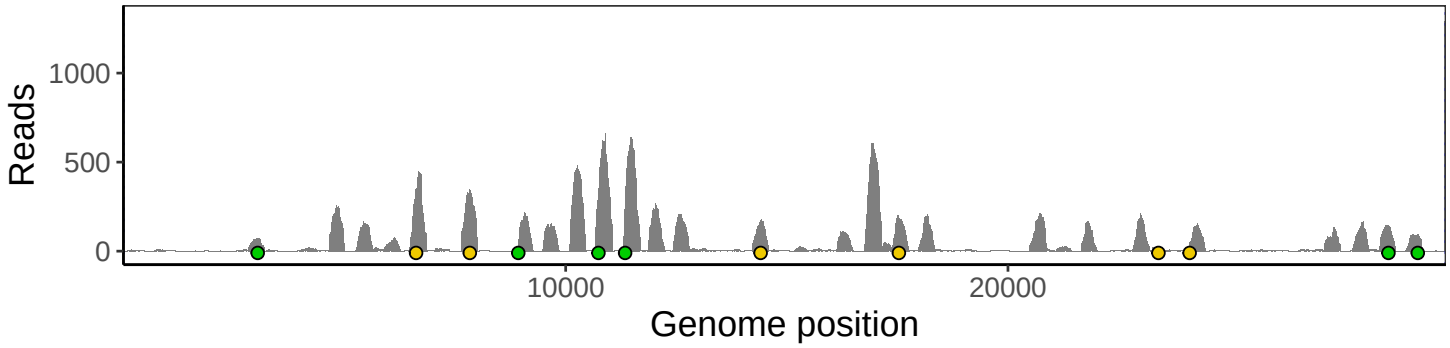

Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



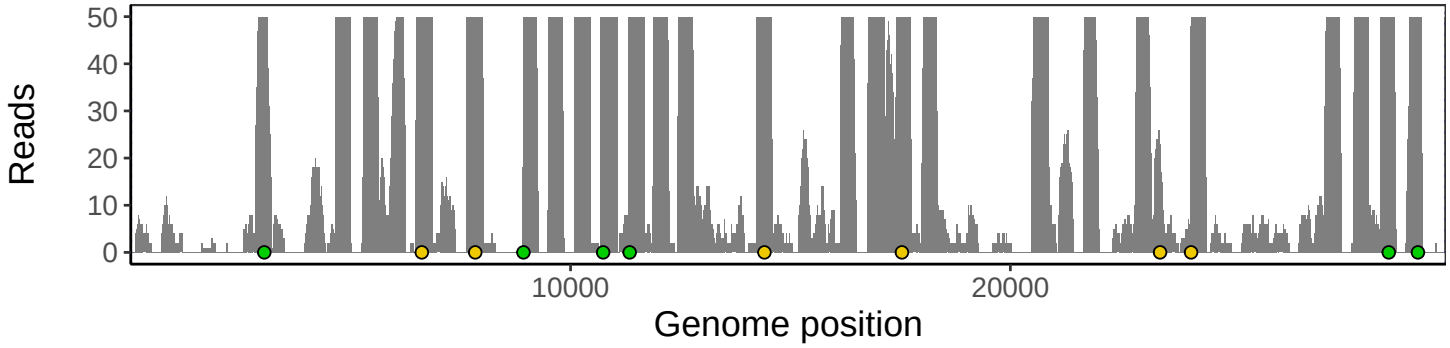

The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

**VSP0778-1 | 2021-02-01 | Saliva | PQ-Seq9 | genomes | single experiment**

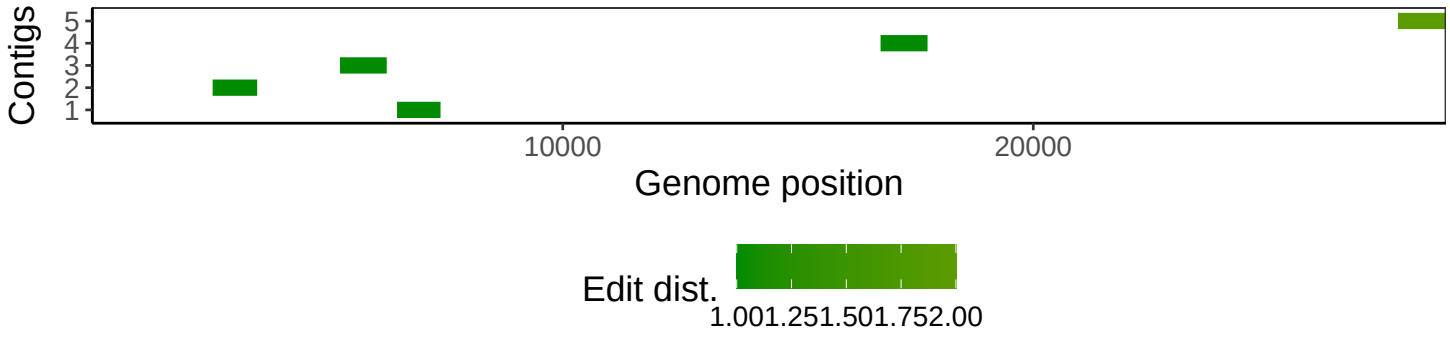
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.




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

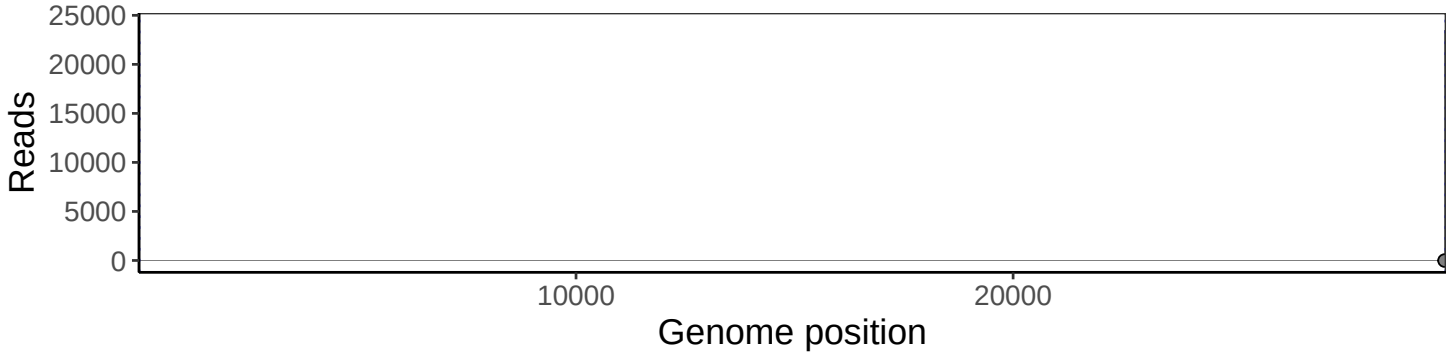



The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

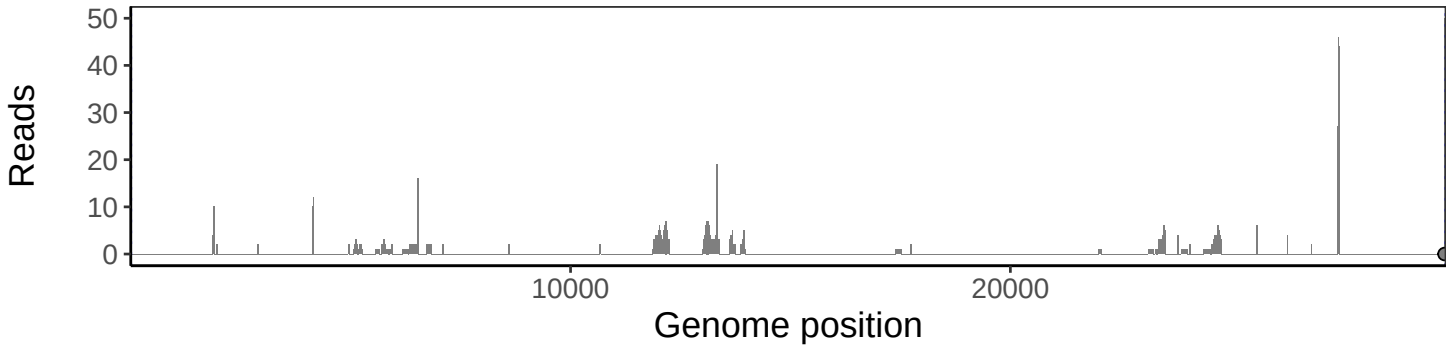


**VSP0778-2 | 2021-02-01 | Saliva | PQ-Seq9 | genomes | single experiment**

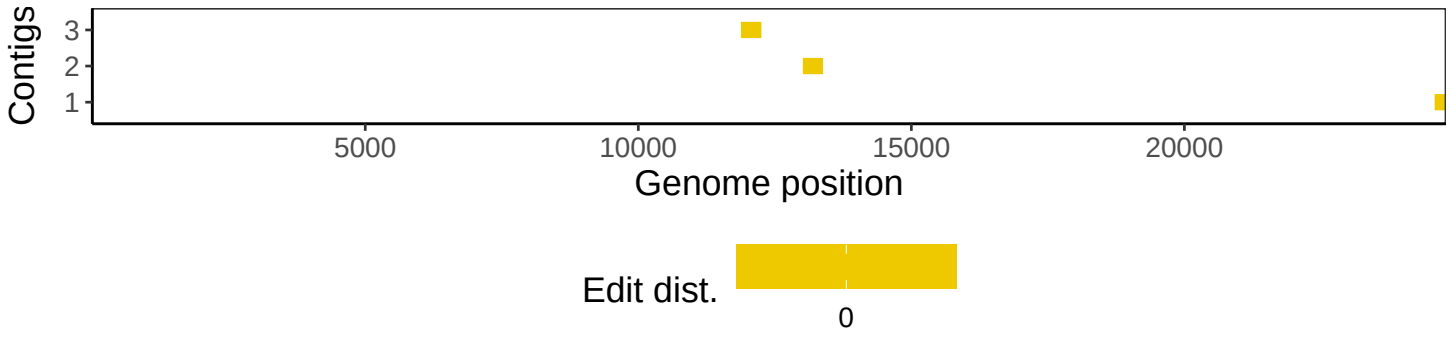
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.




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

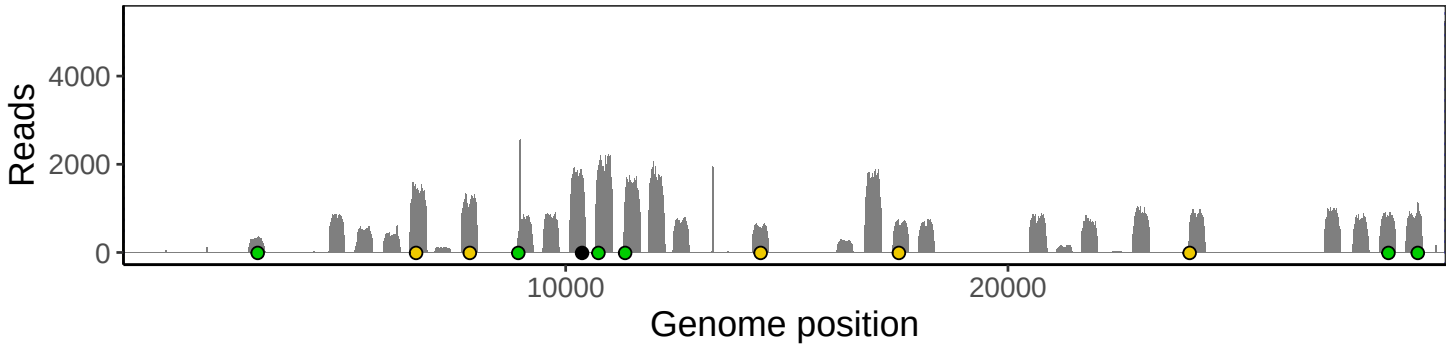



The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

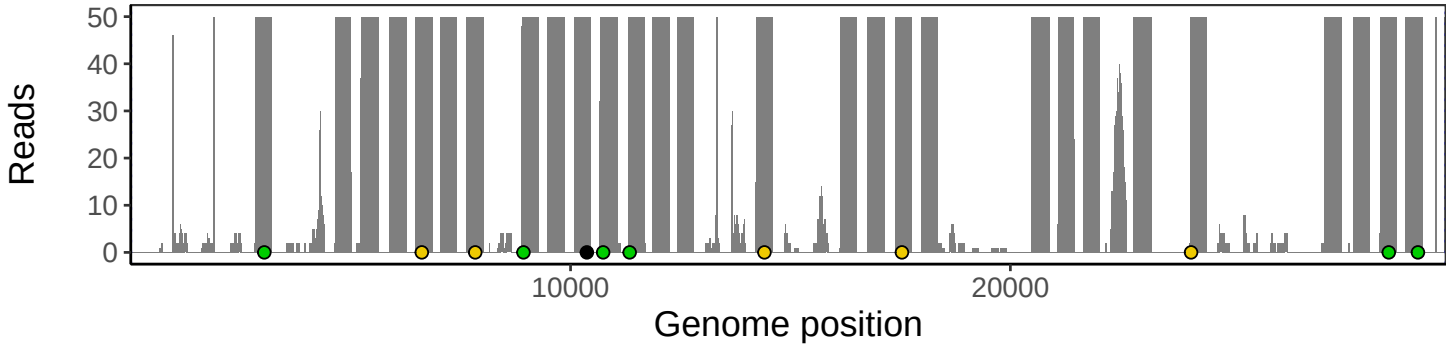


**VSP0778-3 | 2021-02-01 | Saliva | PQ-Seq9 | genomes | single experiment**

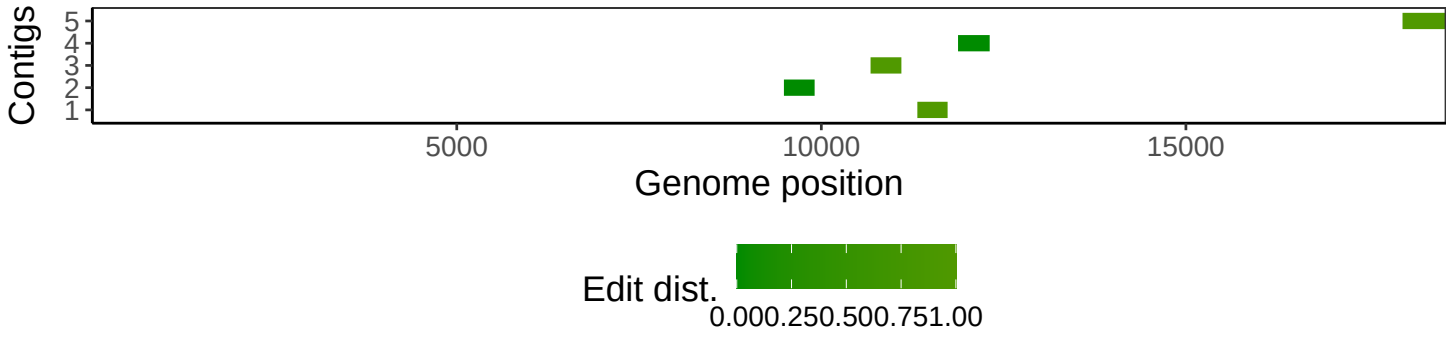
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.




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 | 3.1.3 |
| 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 | 3.3.3 |
| 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 |