COVID-19 subject PQ-Seq6

2021-06-23

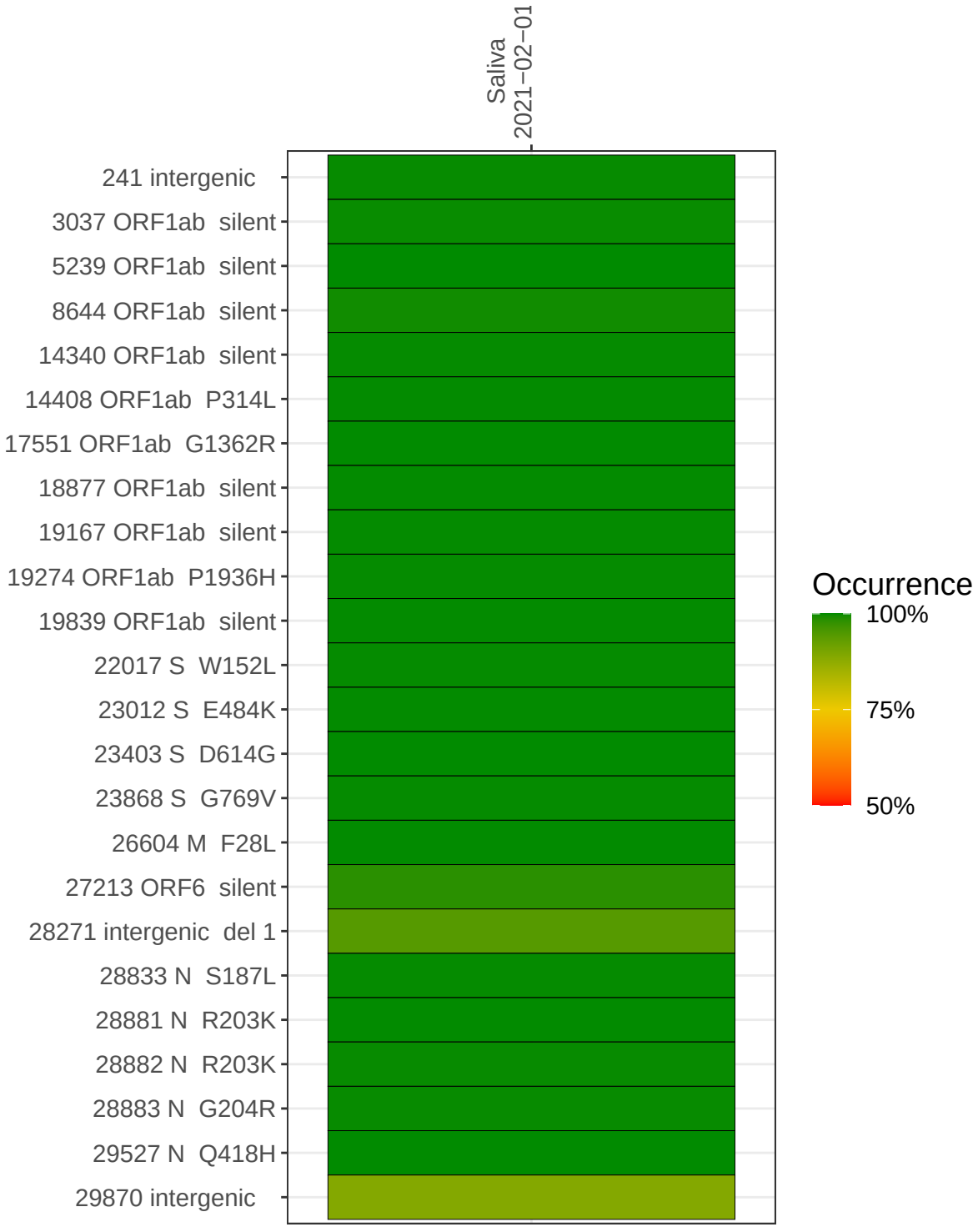
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

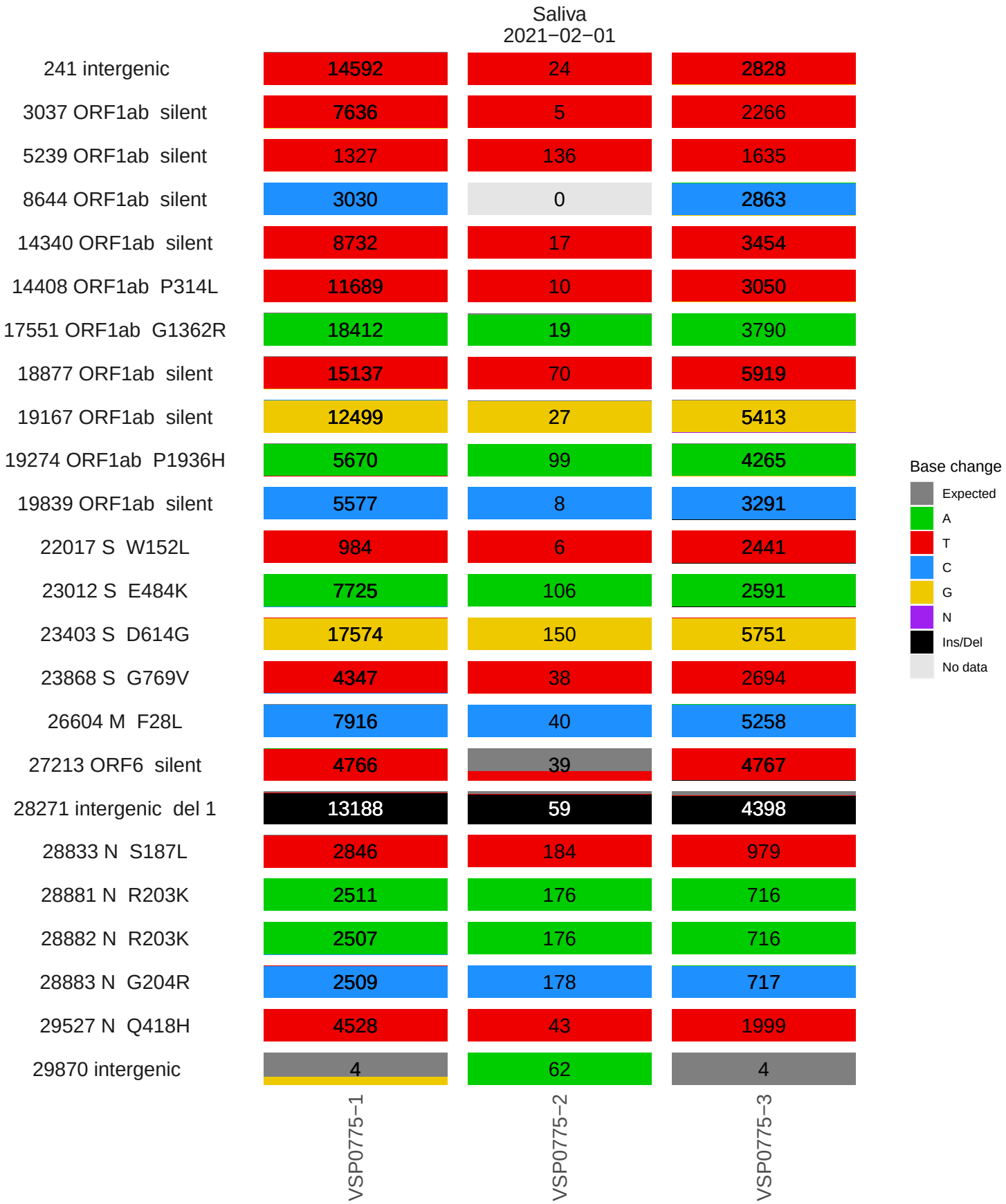
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0775	composite	NA	Saliva	2021-02-01	21.73	R.1	99.6%	99.2%
VSP0775-1	single experiment	NA	Saliva	2021-02-01	21.70	R.1	99.5%	99.0%
VSP0775-2	single experiment	NA	Saliva	2021-02-01	6.18	NA	93.6%	83.1%
VSP0775-3	single experiment	NA	Saliva	2021-02-01	21.73	R.1	99.2%	99.0%

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.

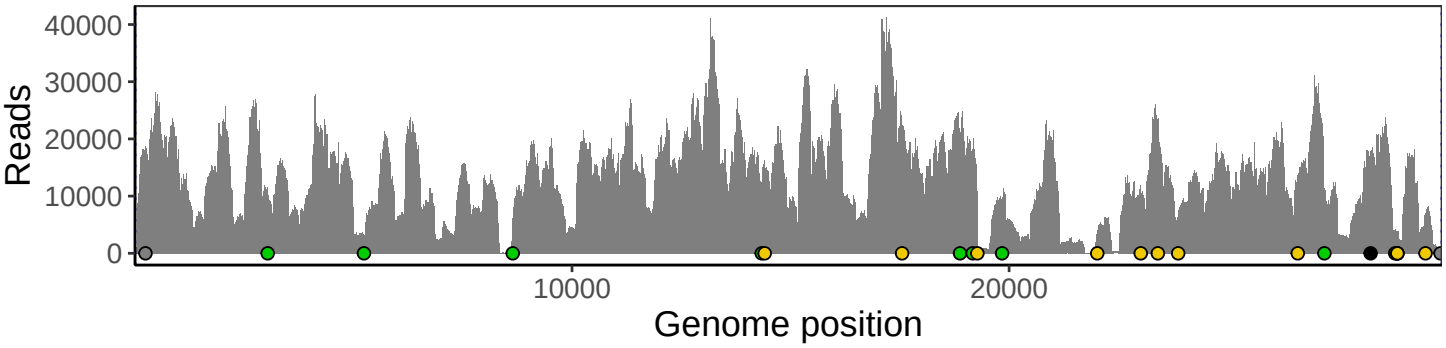



Saliva
2021−02−01
241 intergenic
3037 ORF1ab silent
5239 ORF1ab silent
8644 ORF1ab silent
14340 ORF1ab silent
14408 ORF1ab P314L
17551 ORF1ab G1362R
18877 ORF1ab silent
19167 ORF1ab silent
19274 ORF1ab P1936H
19839 ORF1ab silent
22017 S W152L
23012 S E484K
23403 S D614G
23868 S G769V
26604 M F28L
27213 ORF6 silent
28271 intergenic del 1
28833 N S187L
28881 N R203K
28882 N R203K
28883 N G204R
29527 N Q418H
29870 intergenic
14592
24
2828
7636
5
2266
1327
136
1635
3030
0
2863
8732
17
3454
11689
10
3050
18412
19
3790
15137
70
5919
12499
27
5413
5670
99
4265
5577
8
3291
984
6
2441
7725
106
2591
17574
150
5751
4347
38
2694
7916
40
5258
4766
39
4767
13188
59
4398
2846
184
979
2511
176
716
2507
176
716
2509
178
717
4528
43
1999
4
62
4
VSP0775−1
VSP0775−2
VSP0775−3
Base change
Expected
A
T
C
G
N
Ins/Del
No data

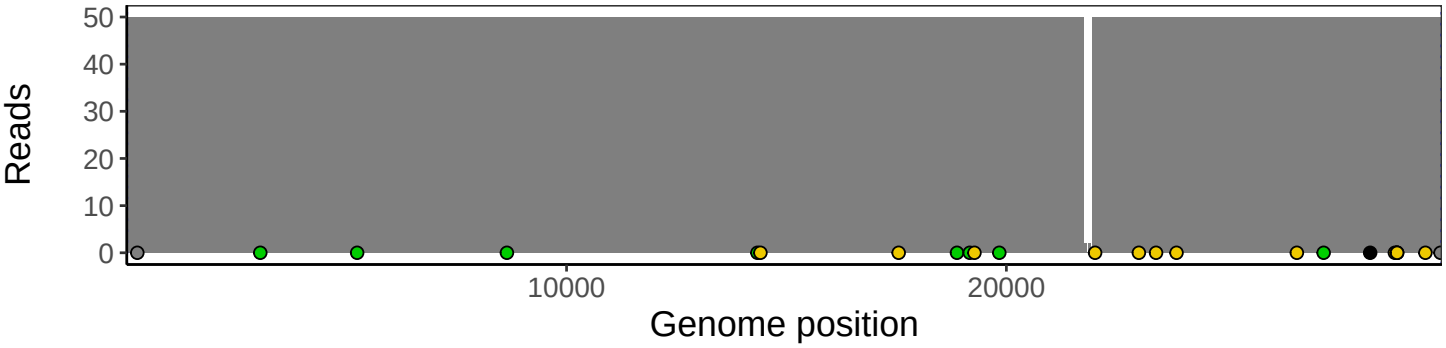
Analyses of individual experiments and composite results

VSP0775 | 2021-02-01 | Saliva | PQ-Seq6 | composite result

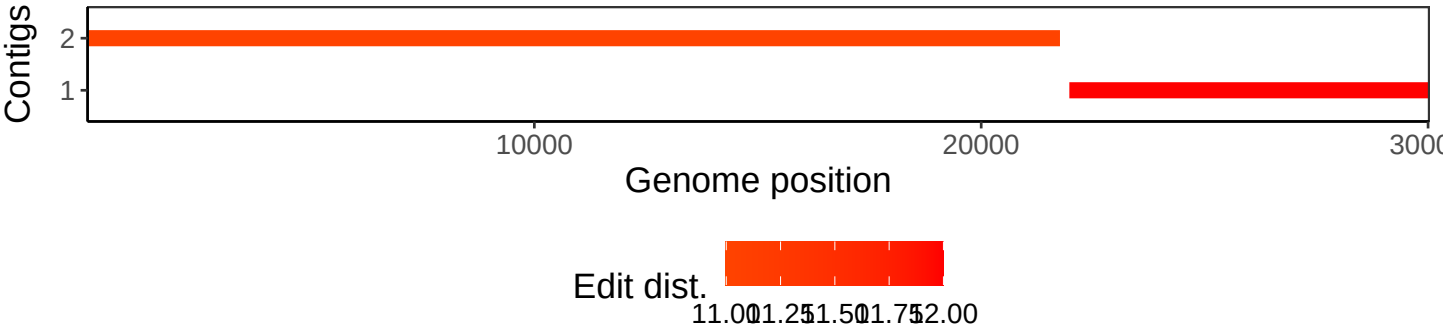
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

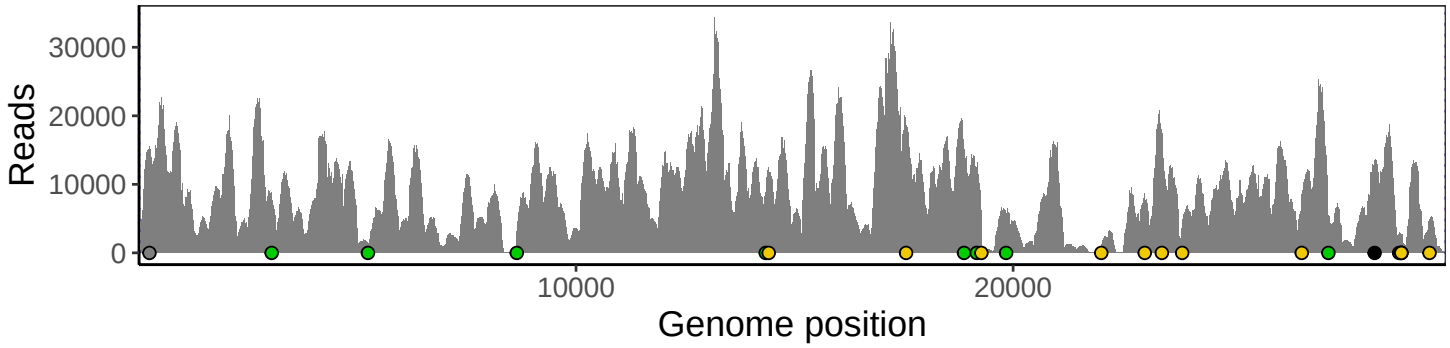


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

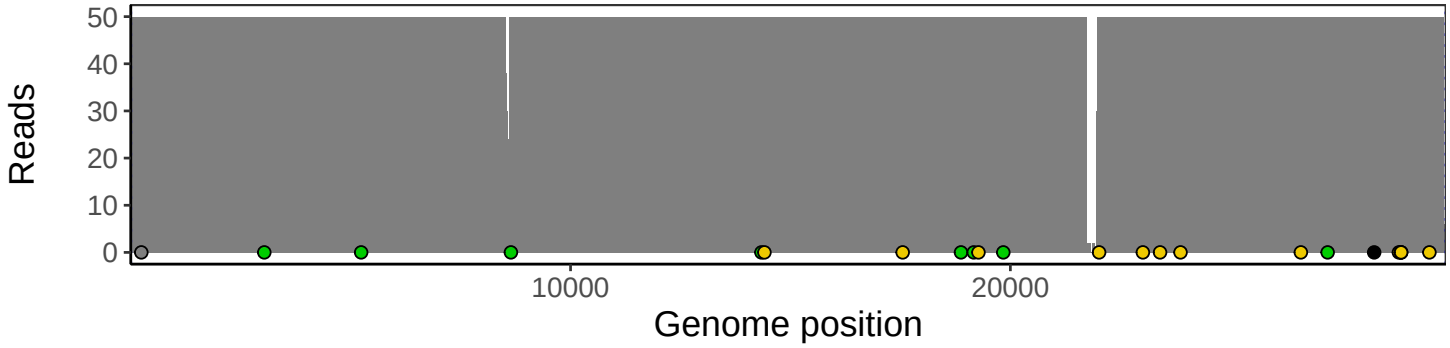


VSP0775-1 | 2021-02-01 | Saliva | PQ-Seq6 | genomes | single experiment

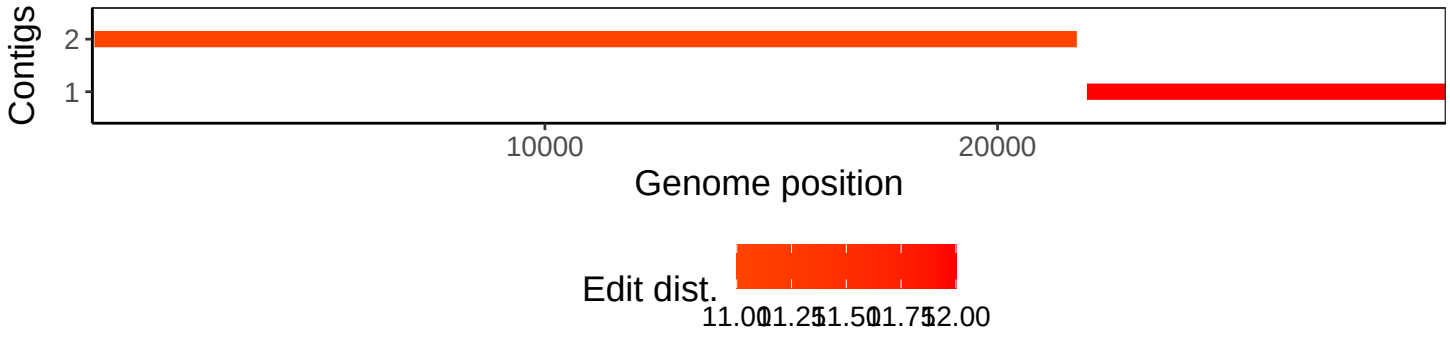
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

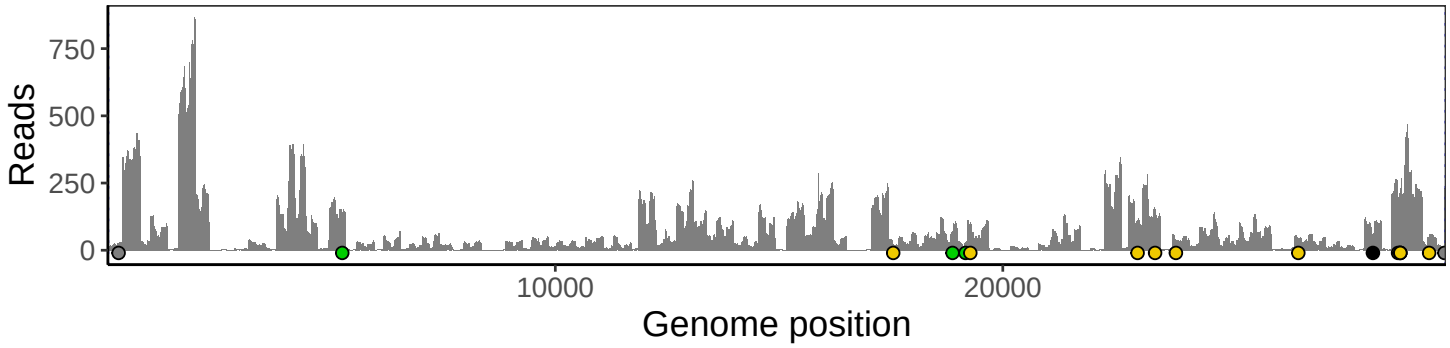


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

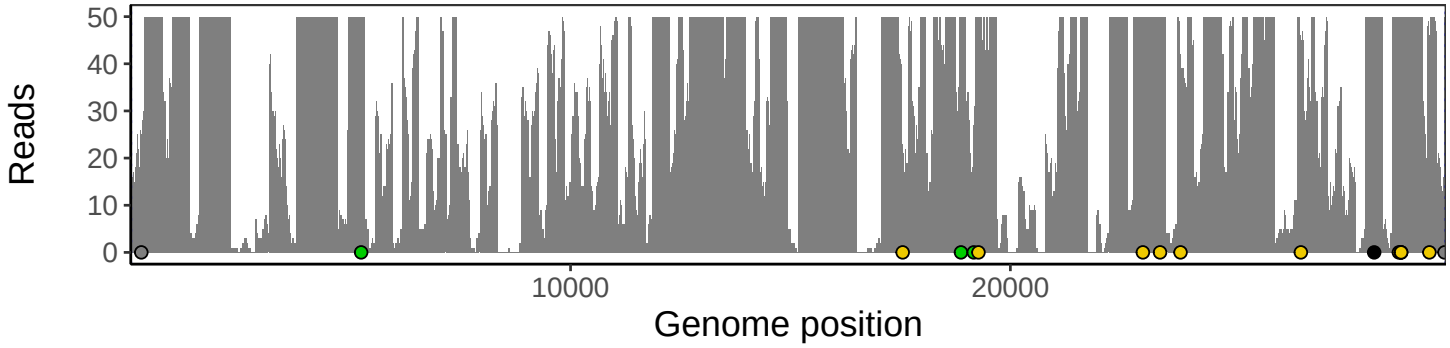


VSP0775-2 | 2021-02-01 | Saliva | PQ-Seq6 | genomes | single experiment

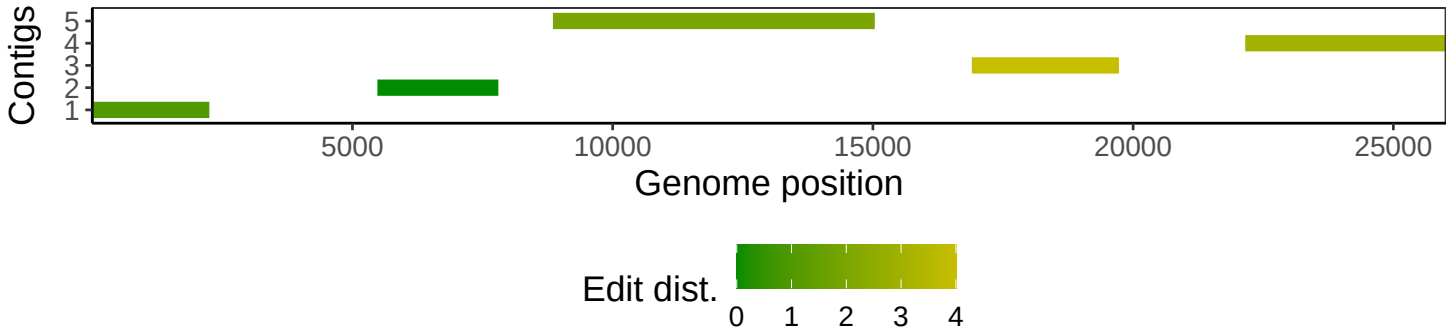
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

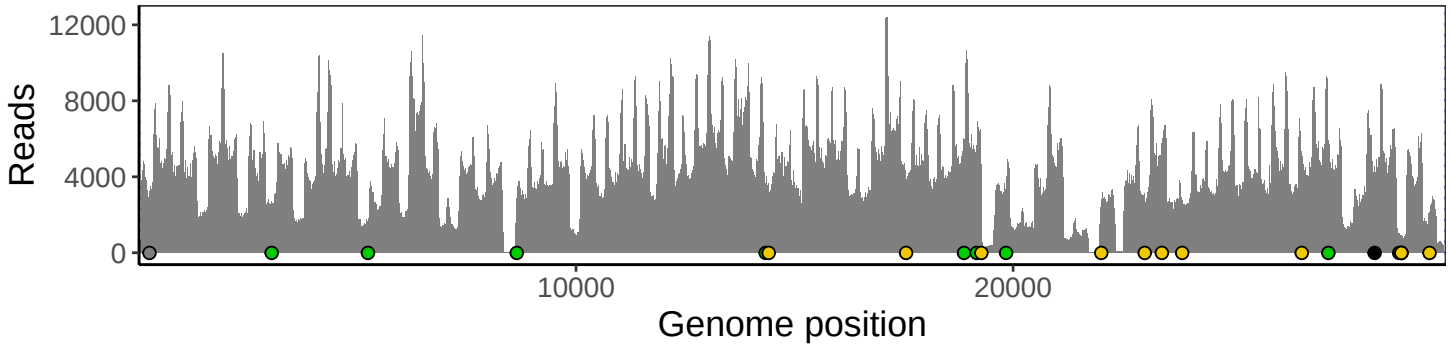


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

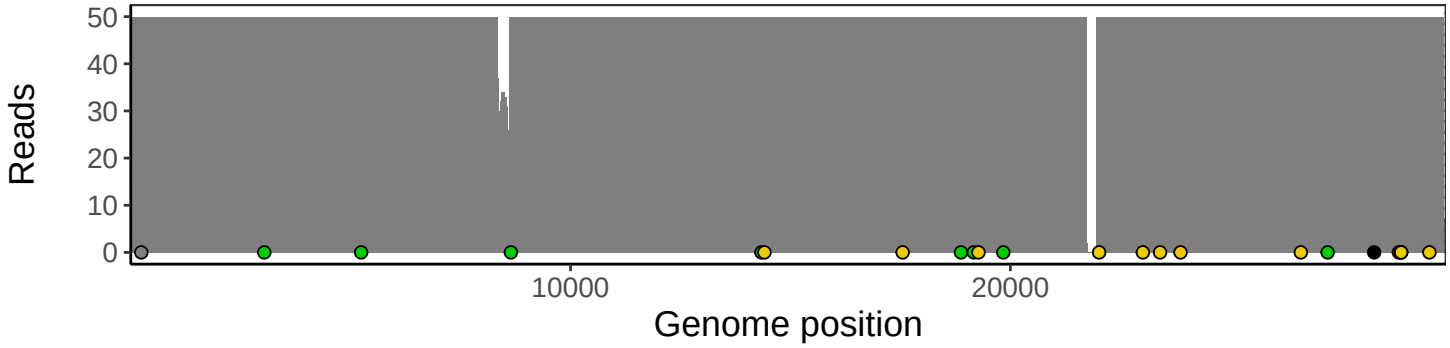


VSP0775-3 | 2021-02-01 | Saliva | PQ-Seq6 | genomes | single experiment

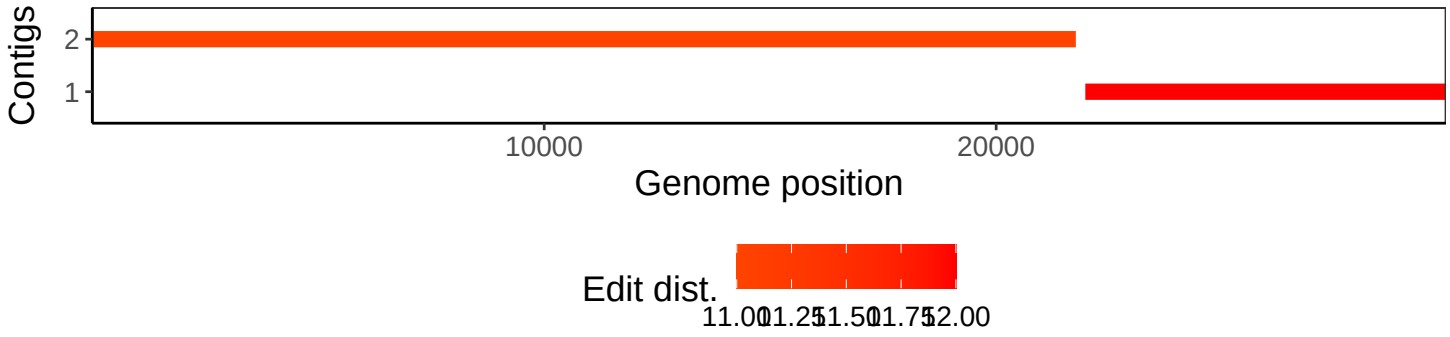
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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