COVID-19 subject PQ-Seq4

2021-04-17

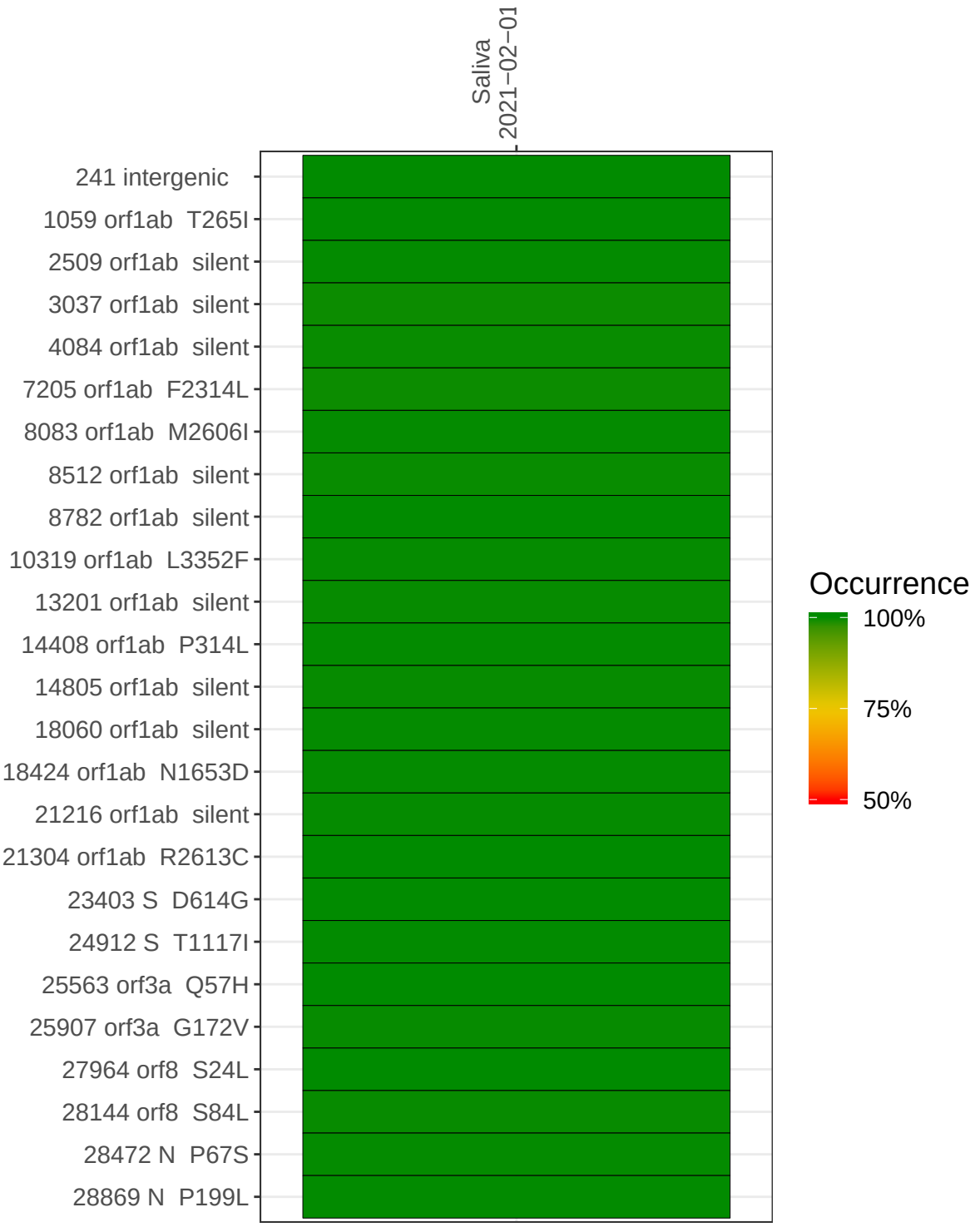
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

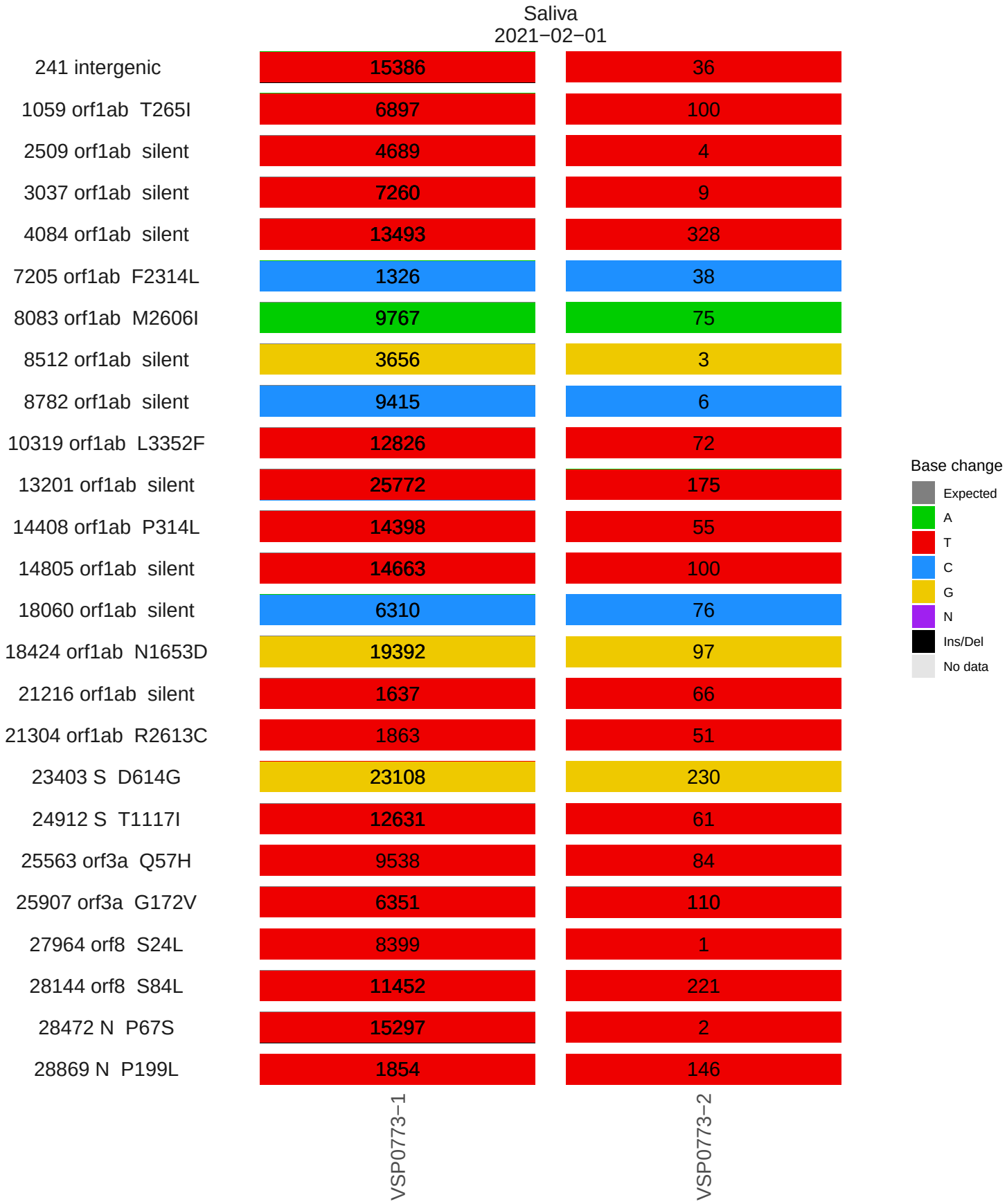
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0773	composite	NA	Saliva	2021-02-01	29.82	B.1.2	99.8%	99.8%
VSP0773-1	single experiment	NA	Saliva	2021-02-01	29.84	B.1.2	99.8%	99.8%
VSP0773-2	single experiment	NA	Saliva	2021-02-01	7.91	NA	97.6%	90.3%

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.

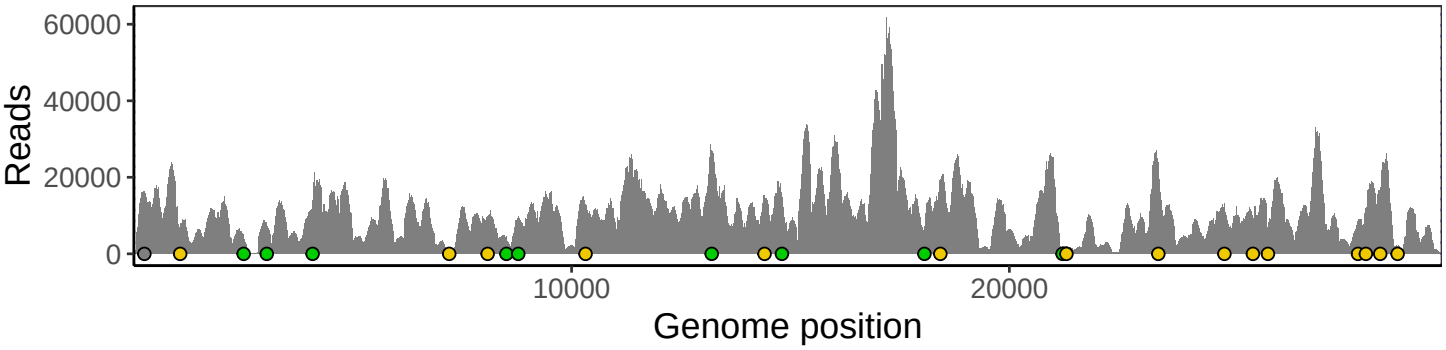



Saliva
2021−02−01
241 intergenic
1059 orf1ab T265I
2509 orf1ab silent
3037 orf1ab silent
4084 orf1ab silent
7205 orf1ab F2314L
8083 orf1ab M2606I
8512 orf1ab silent
8782 orf1ab silent
10319 orf1ab L3352F
13201 orf1ab silent
14408 orf1ab P314L
14805 orf1ab silent
18060 orf1ab silent
18424 orf1ab N1653D
21216 orf1ab silent
21304 orf1ab R2613C
23403 S D614G
24912 S T1117I
25563 orf3a Q57H
25907 orf3a G172V
27964 orf8 S24L
28144 orf8 S84L
28472 N P67S
28869 N P199L
15386
6897
4689
7260
13493
1326
9767
3656
9415
12826
25772
14398
14663
6310
19392
1637
1863
23108
12631
9538
6351
8399
11452
15297
1854
36
100
4
9
328
38
75
3
6
72
175
55
100
76
97
66
51
230
61
84
110
1
221
2
146
VSP0773−1
VSP0773−2
Base change
Expected
A
T
C
G
N
Ins/Del
No data

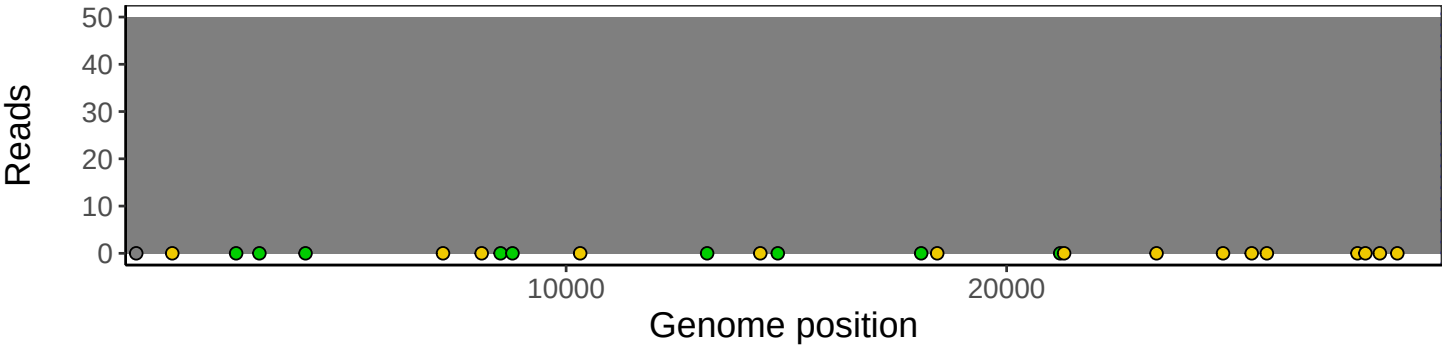
Analyses of individual experiments and composite results

VSP0773 | 2021-02-01 | Saliva | PQ-Seq4 | composite result

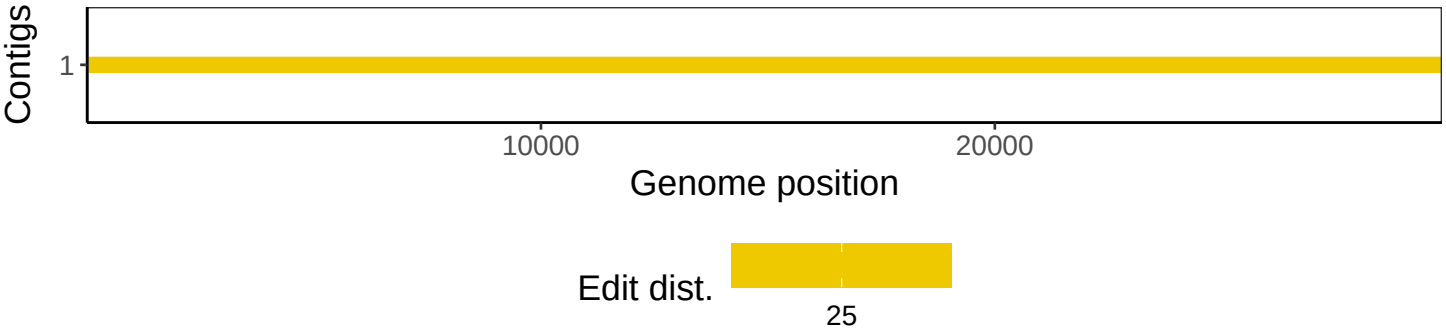
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

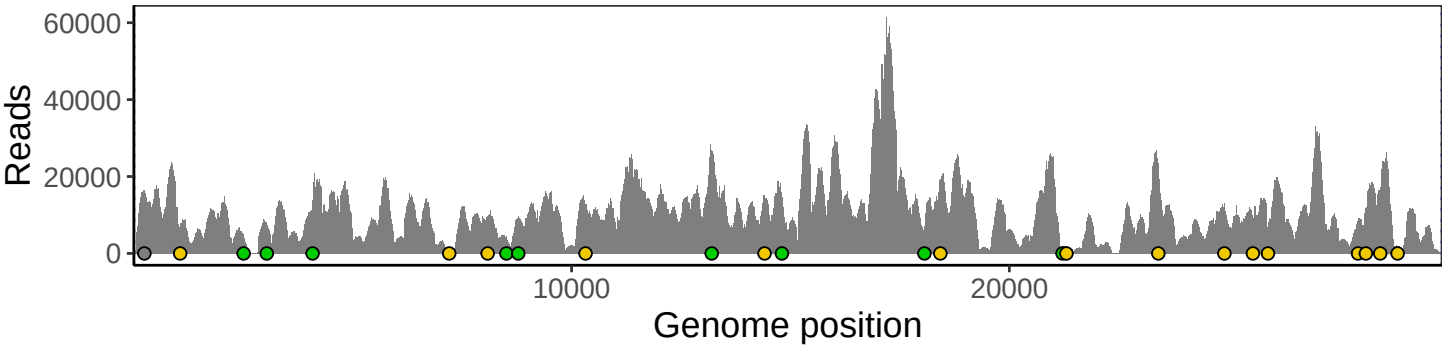


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

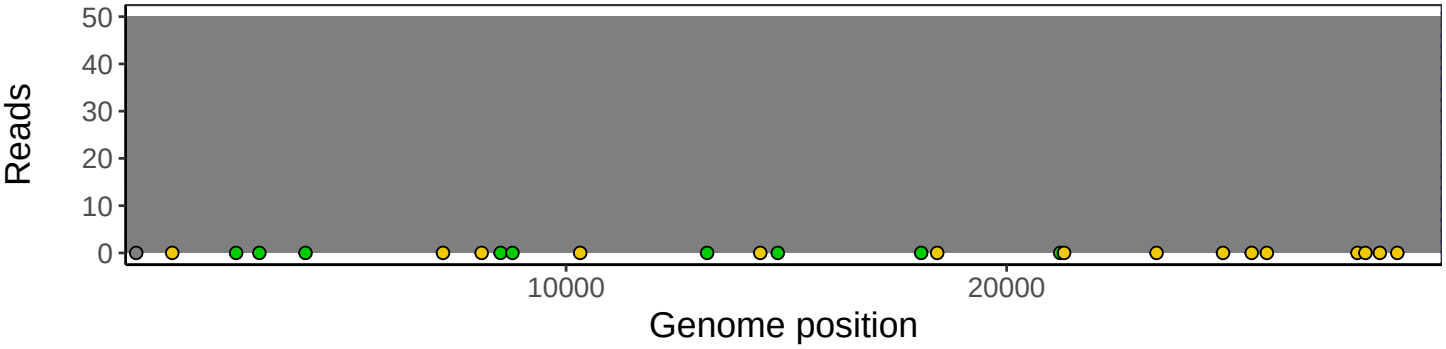


VSP0773-1 | 2021-02-01 | Saliva | PQ-Seq4 | genomes | single experiment

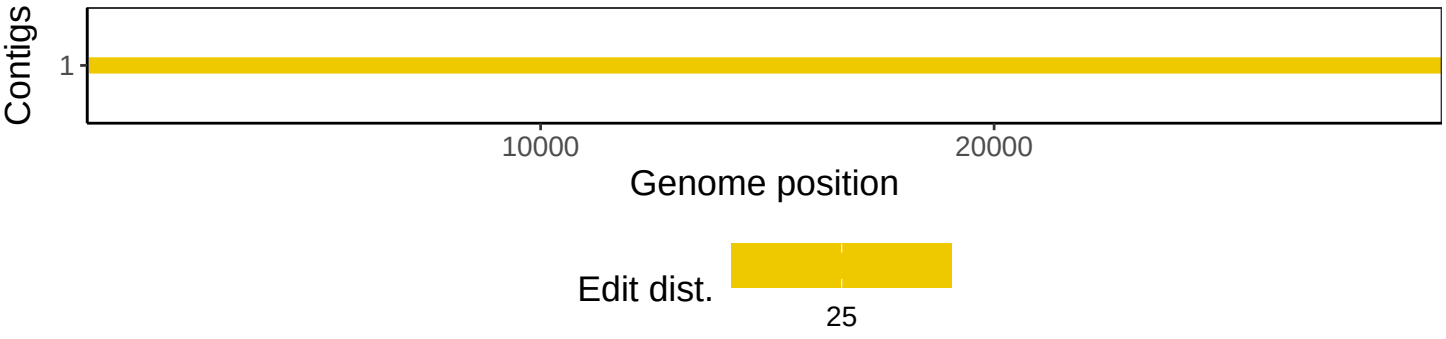
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

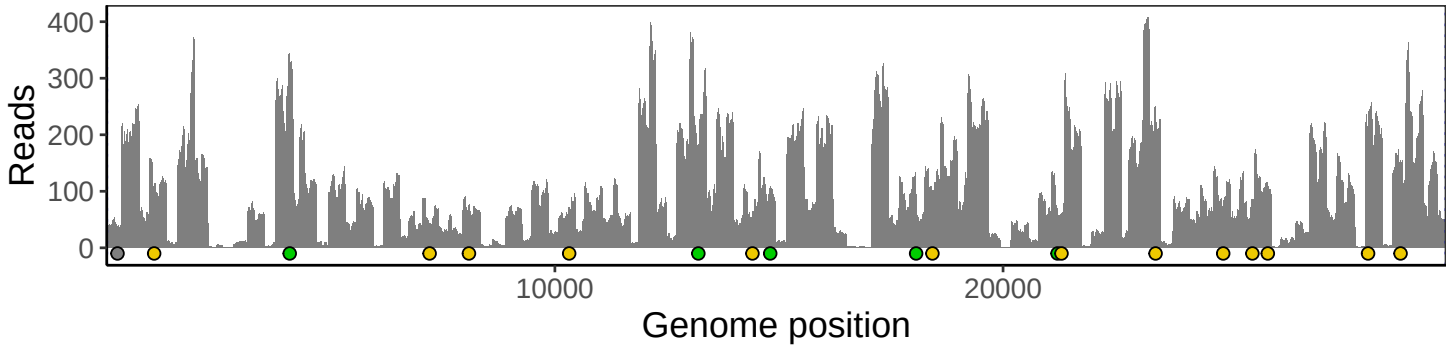


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

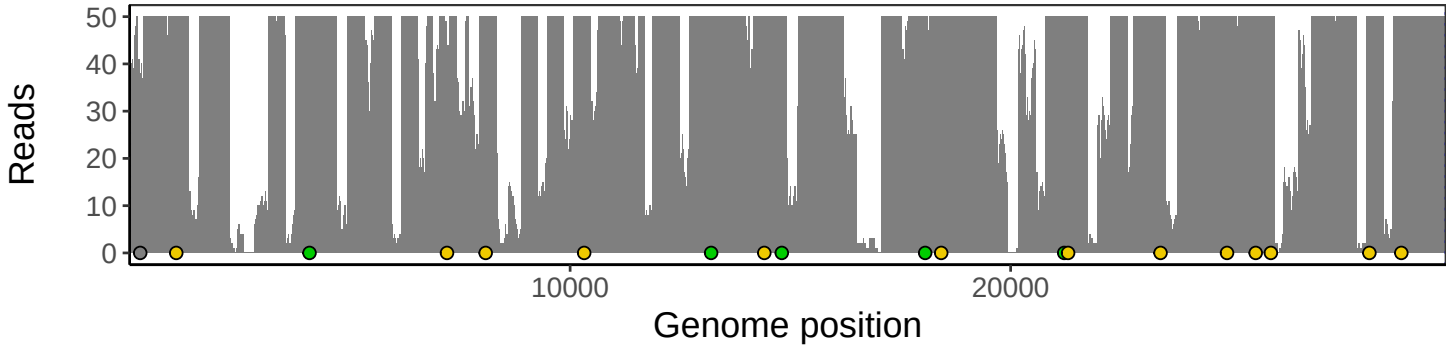


VSP0773-2 | 2021-02-01 | Saliva | PQ-Seq4 | genomes | single experiment

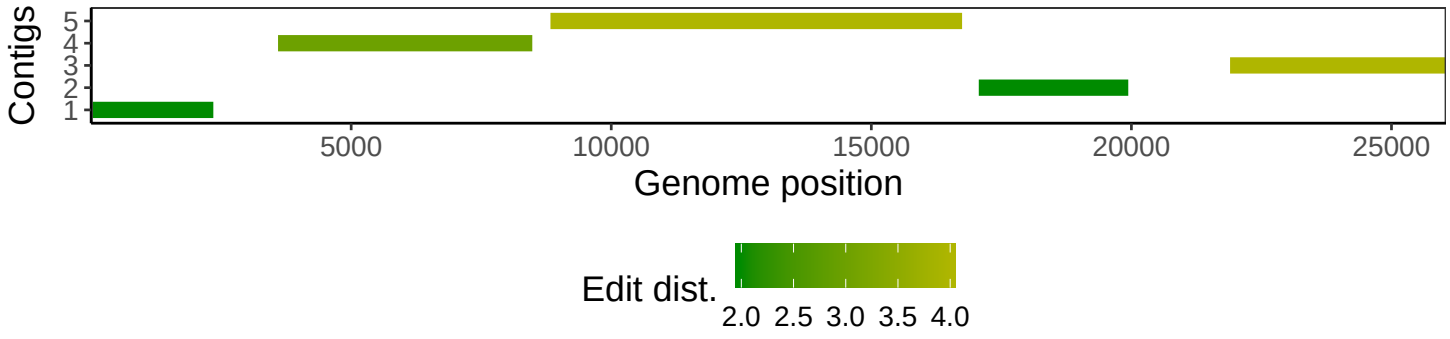
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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