COVID-19 subject PQ-Seq10

2021-05-05

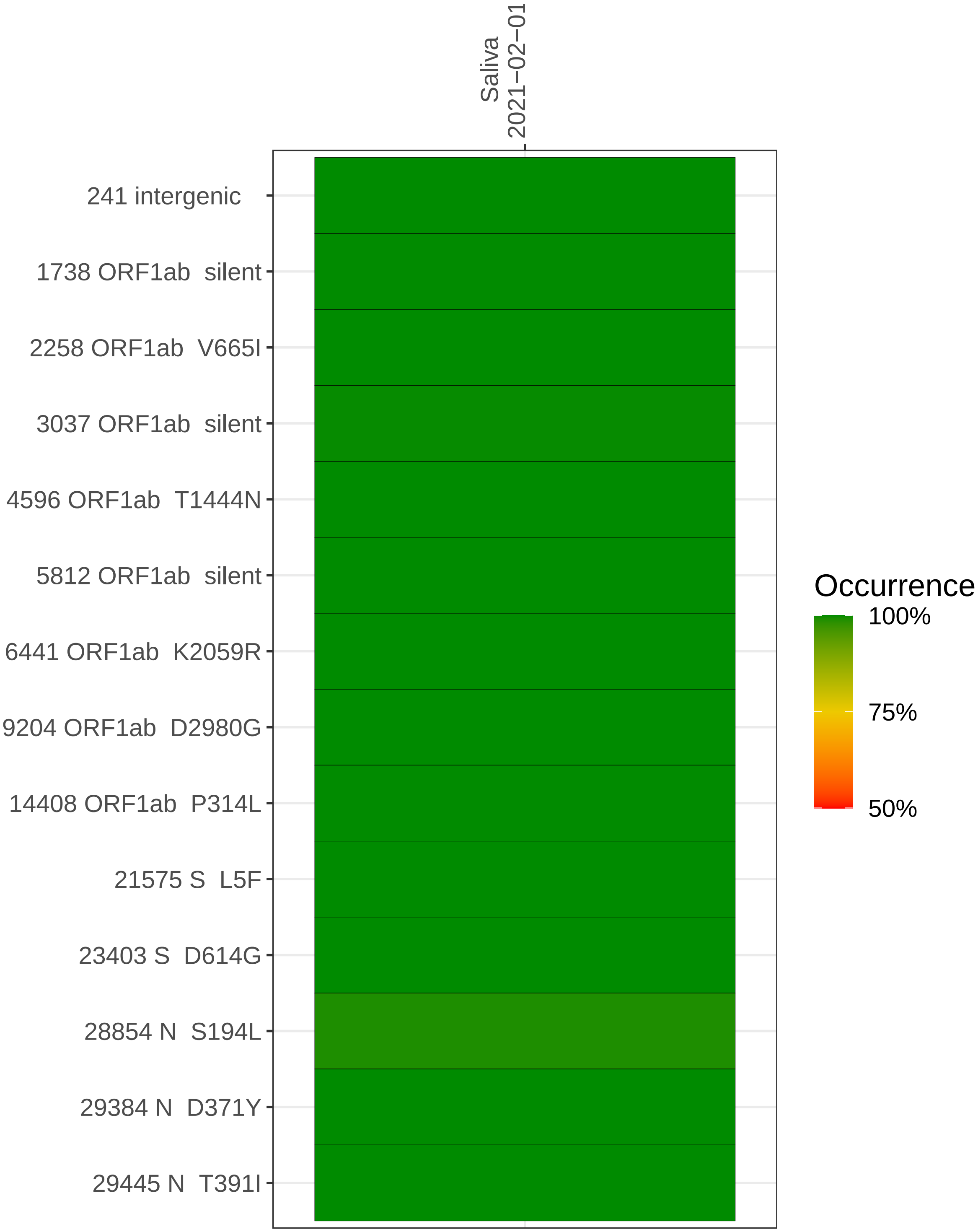
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

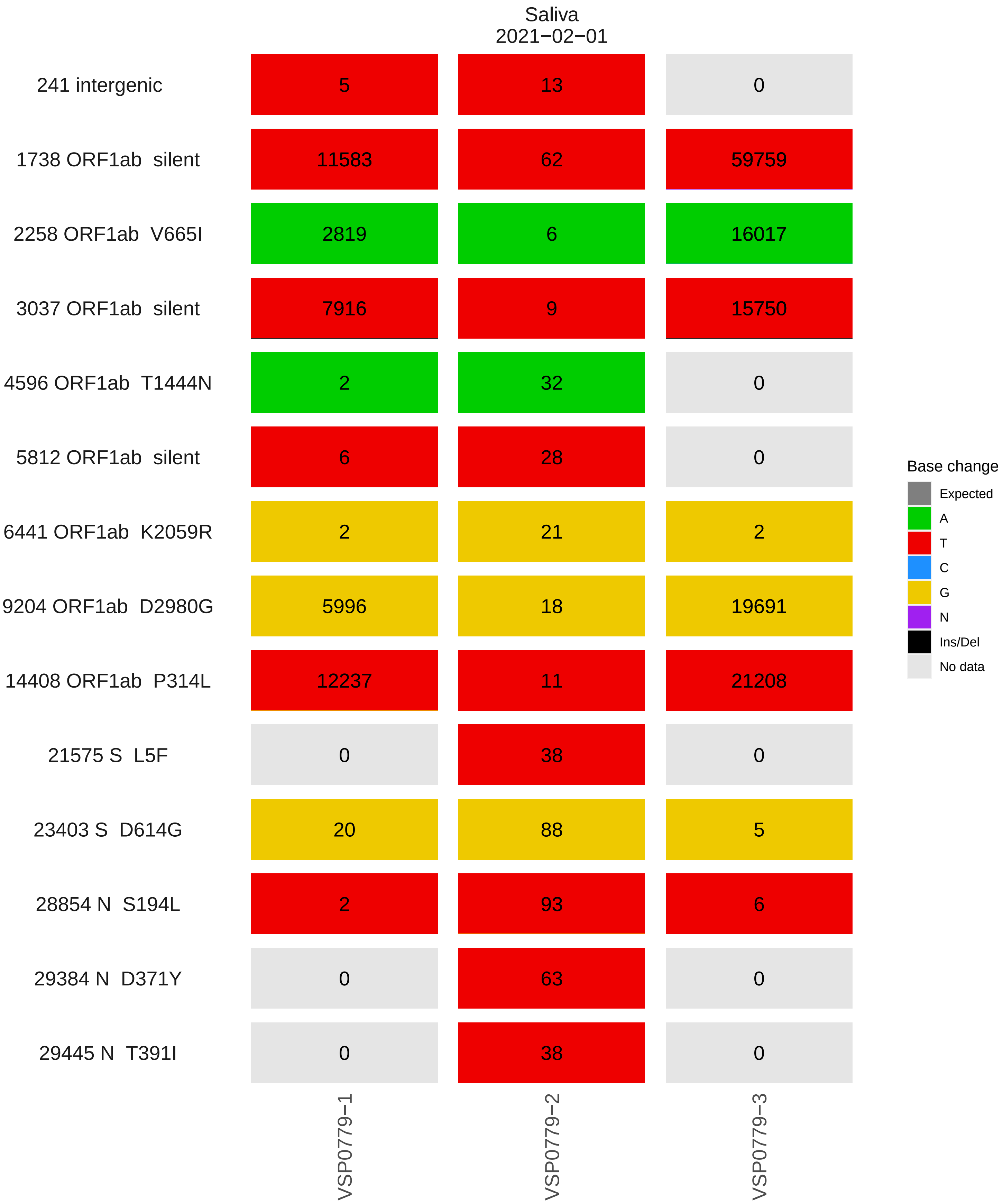
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0779	composite	NA	Saliva	2021-02-01	9.23	B.1.234	99.9%	97.3%
VSP0779-1	single experiment	NA	Saliva	2021-02-01	1.72	NA	90.3%	75.7%
VSP0779-2	single experiment	NA	Saliva	2021-02-01	6.49	NA	96.0%	85.2%
VSP0779-3	single experiment	NA	Saliva	2021-02-01	1.12	NA	70.4%	65.4%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/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

	VSP0779−1	VSP0779−2	VSP0779−3
241 intergenic	🟥 5	🟥 13	⬜ 0
1738 ORF1ab silent	🟥 11583	🟥 62	🟥 59759
2258 ORF1ab V665I	🟩 2819	🟩 6	🟩 16017
3037 ORF1ab silent	🟥 7916	🟥 9	🟥 15750
4596 ORF1ab T1444N	🟩 2	🟩 32	⬜ 0
5812 ORF1ab silent	🟥 6	🟥 28	⬜ 0
6441 ORF1ab K2059R	🟨 2	🟨 21	🟨 2
9204 ORF1ab D2980G	🟨 5996	🟨 18	🟨 19691
14408 ORF1ab P314L	🟥 12237	🟥 11	🟥 21208
21575 S L5F	⬜ 0	🟥 38	⬜ 0
23403 S D614G	🟨 20	🟨 88	🟨 5
28854 N S194L	🟥 2	🟥 93	🟥 6
29384 N D371Y	⬜ 0	🟥 63	⬜ 0
29445 N T391I	⬜ 0	🟥 38	⬜ 0

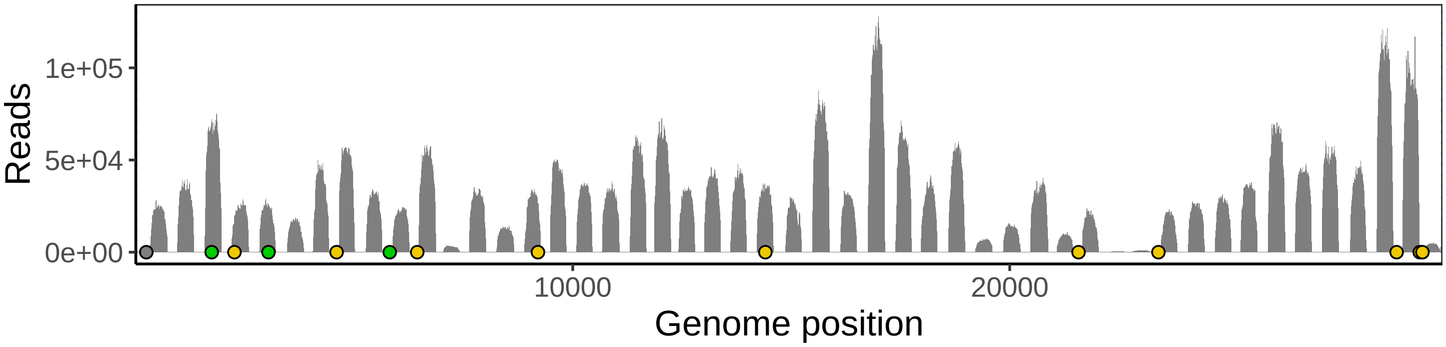
Base change

⬛	Expected
🟩	A
🟥	T
🟦	C
🟨	G
🟪	N
⬛	Ins/Del
⬜	No data

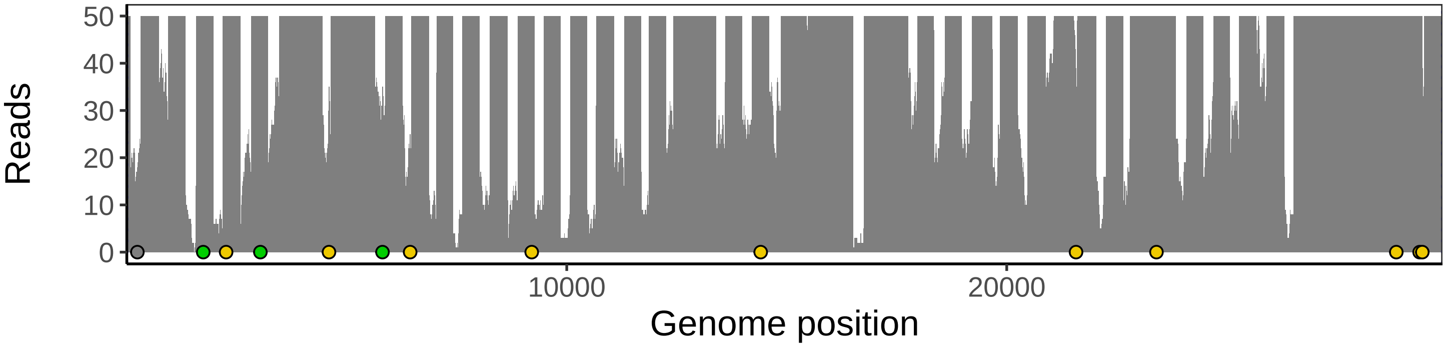
Analyses of individual experiments and composite results

VSP0779 | 2021-02-01 | Saliva | PQ-Seq10 | composite result

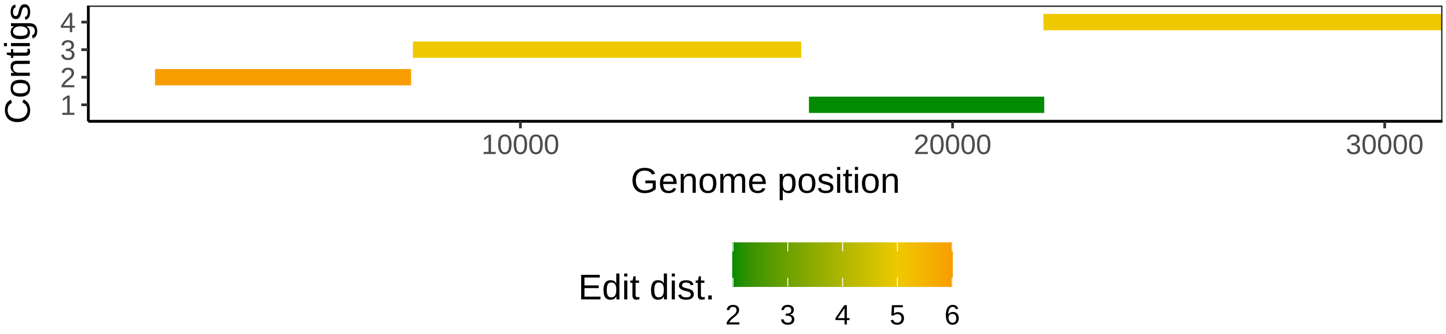
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

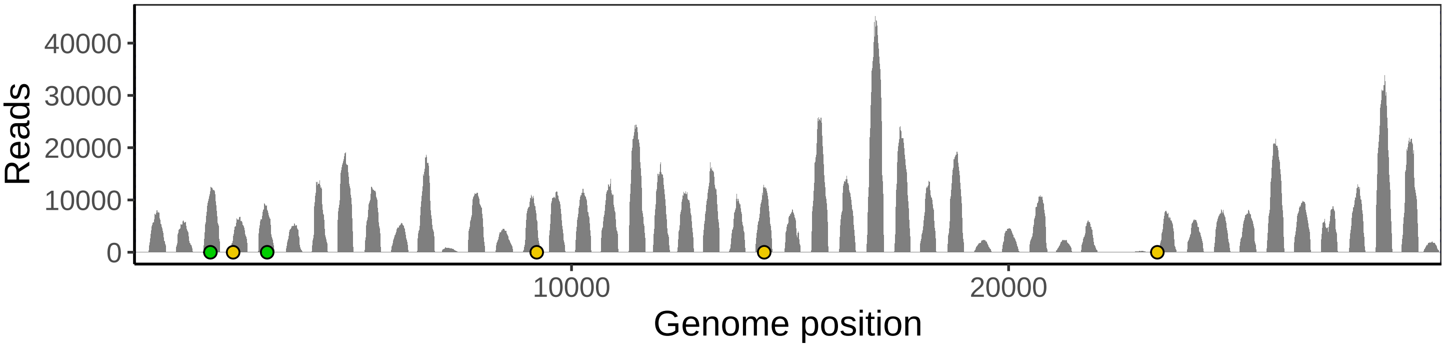


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

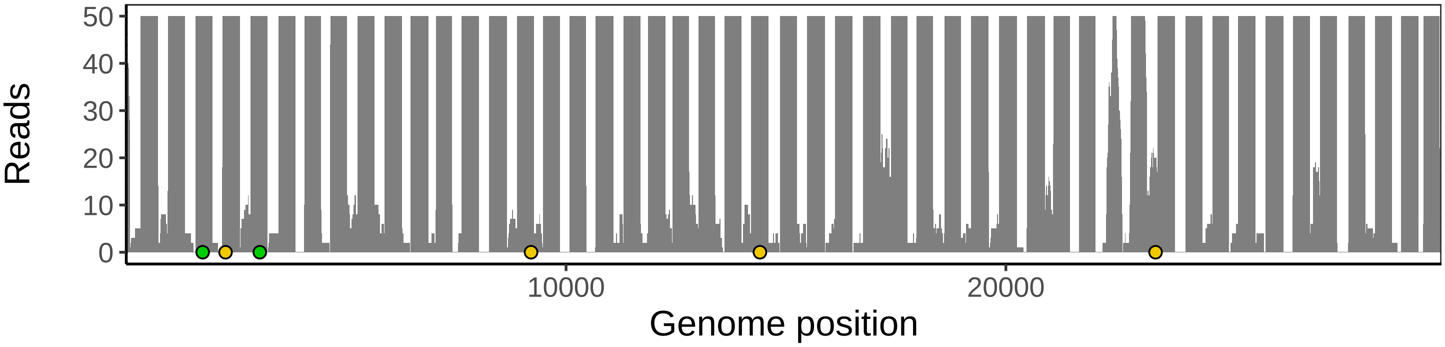


VSP0779-1 | 2021-02-01 | Saliva | PQ-Seq10 | genomes | single experiment

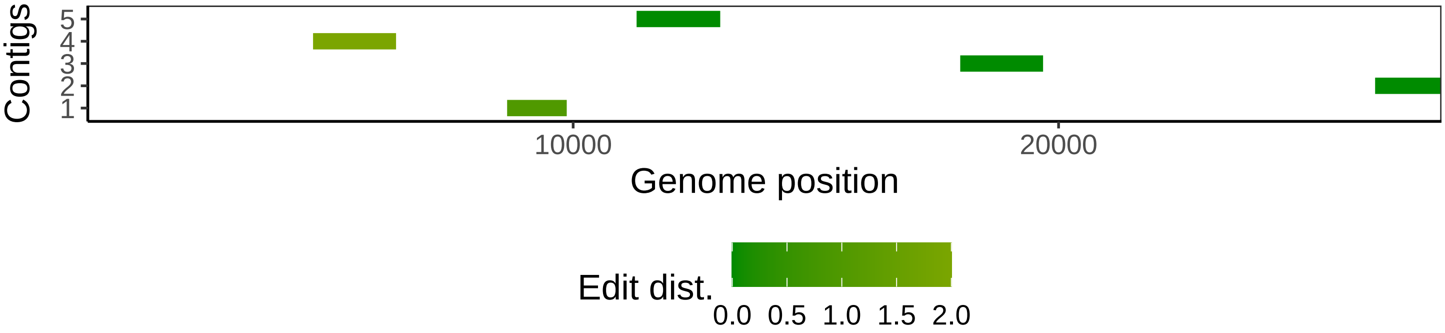
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

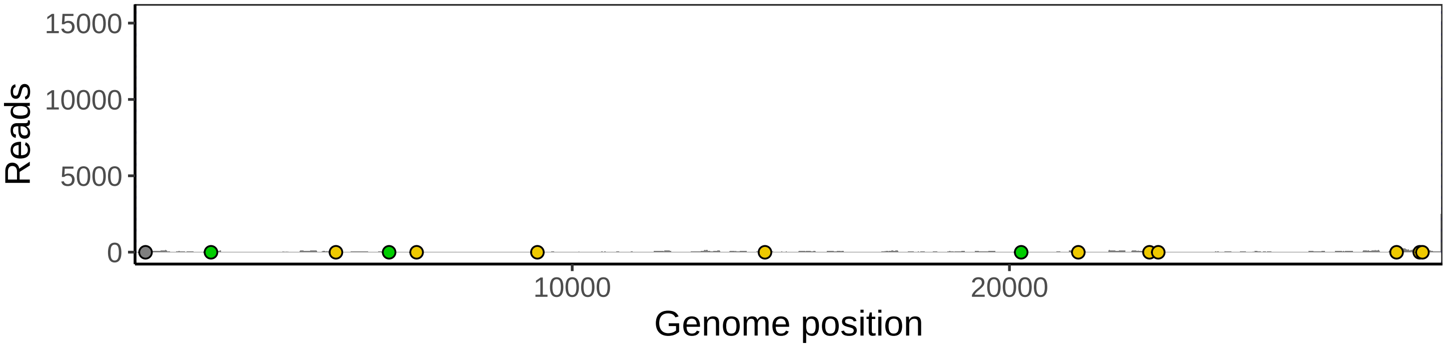


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

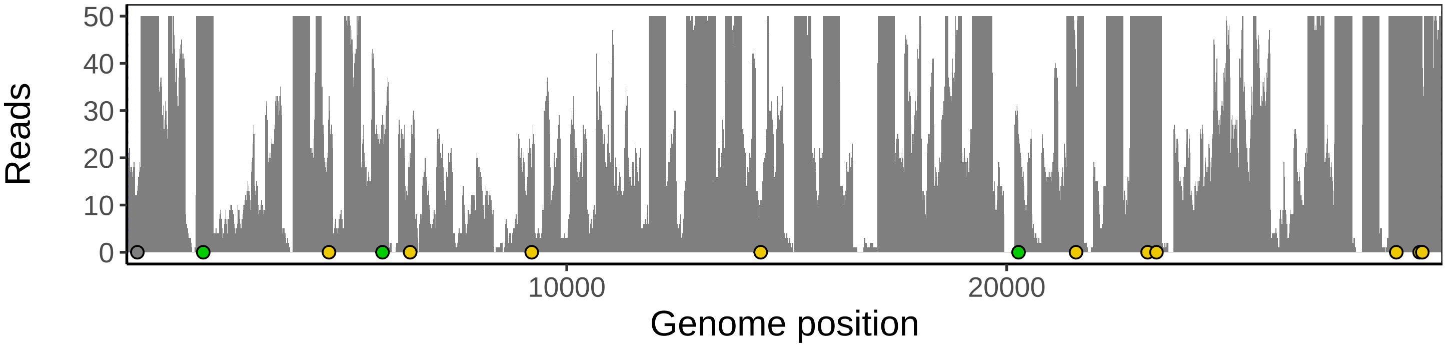


VSP0779-2 | 2021-02-01 | Saliva | PQ-Seq10 | genomes | single experiment

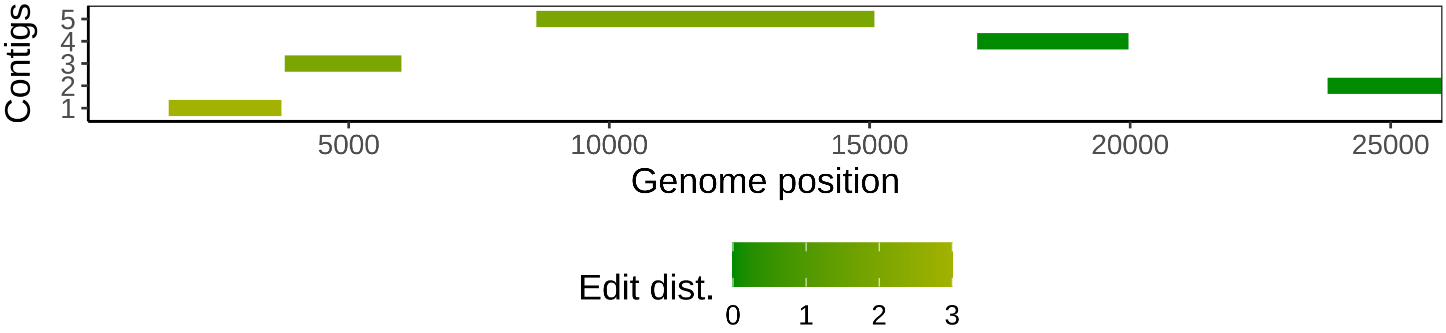
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

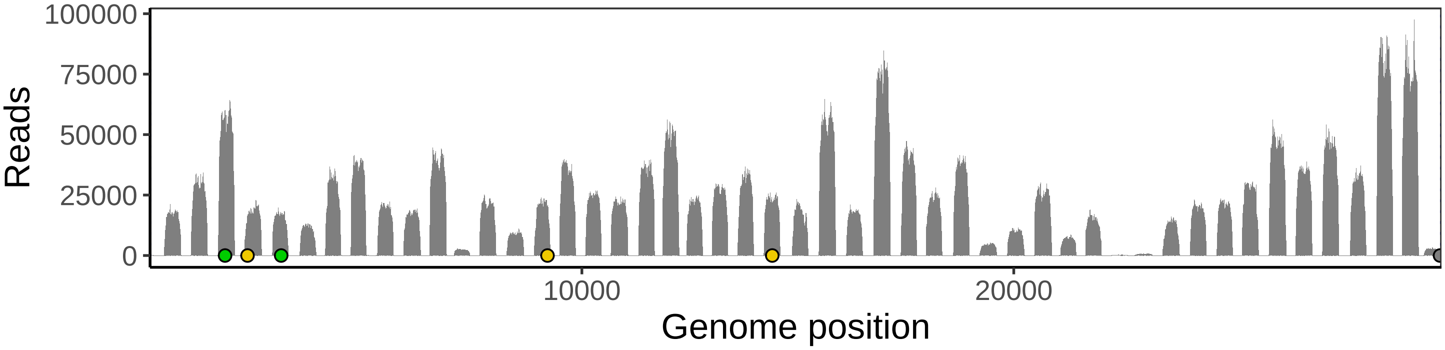


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

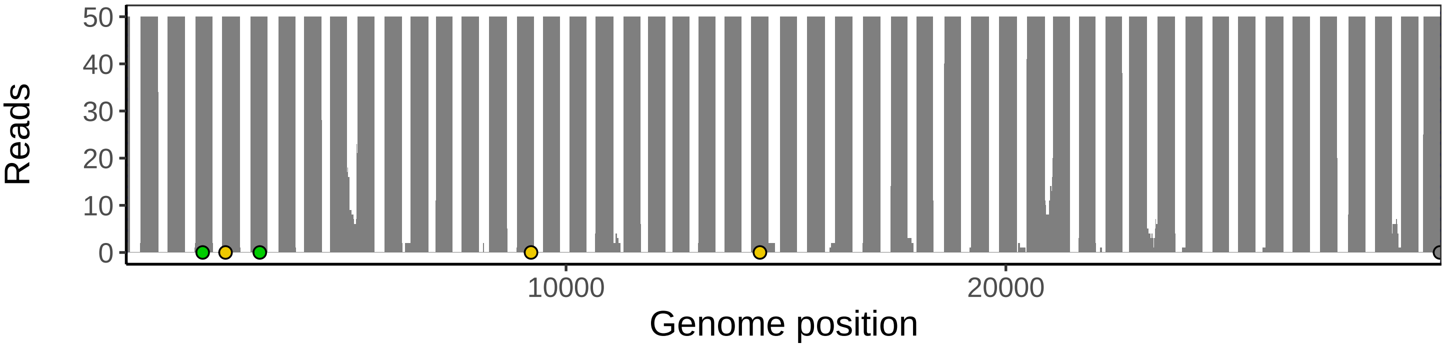


VSP0779-3 | 2021-02-01 | Saliva | PQ-Seq10 | genomes | single experiment

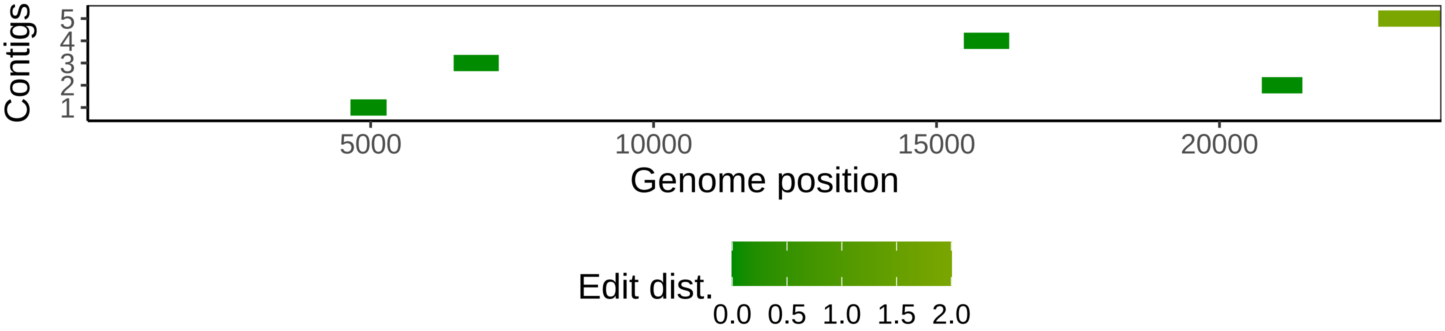
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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