COVID-19 subject 436

2021-03-01

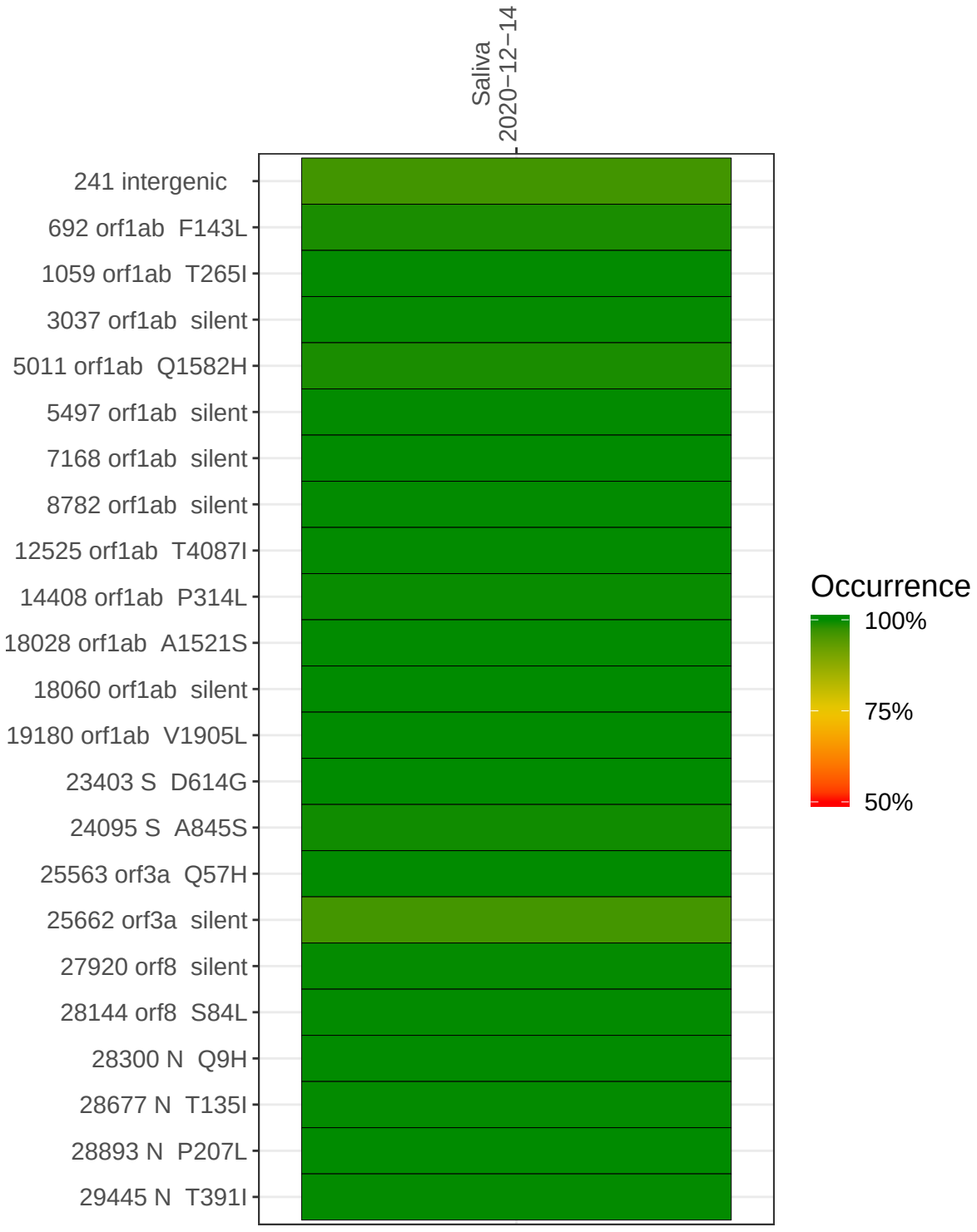
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

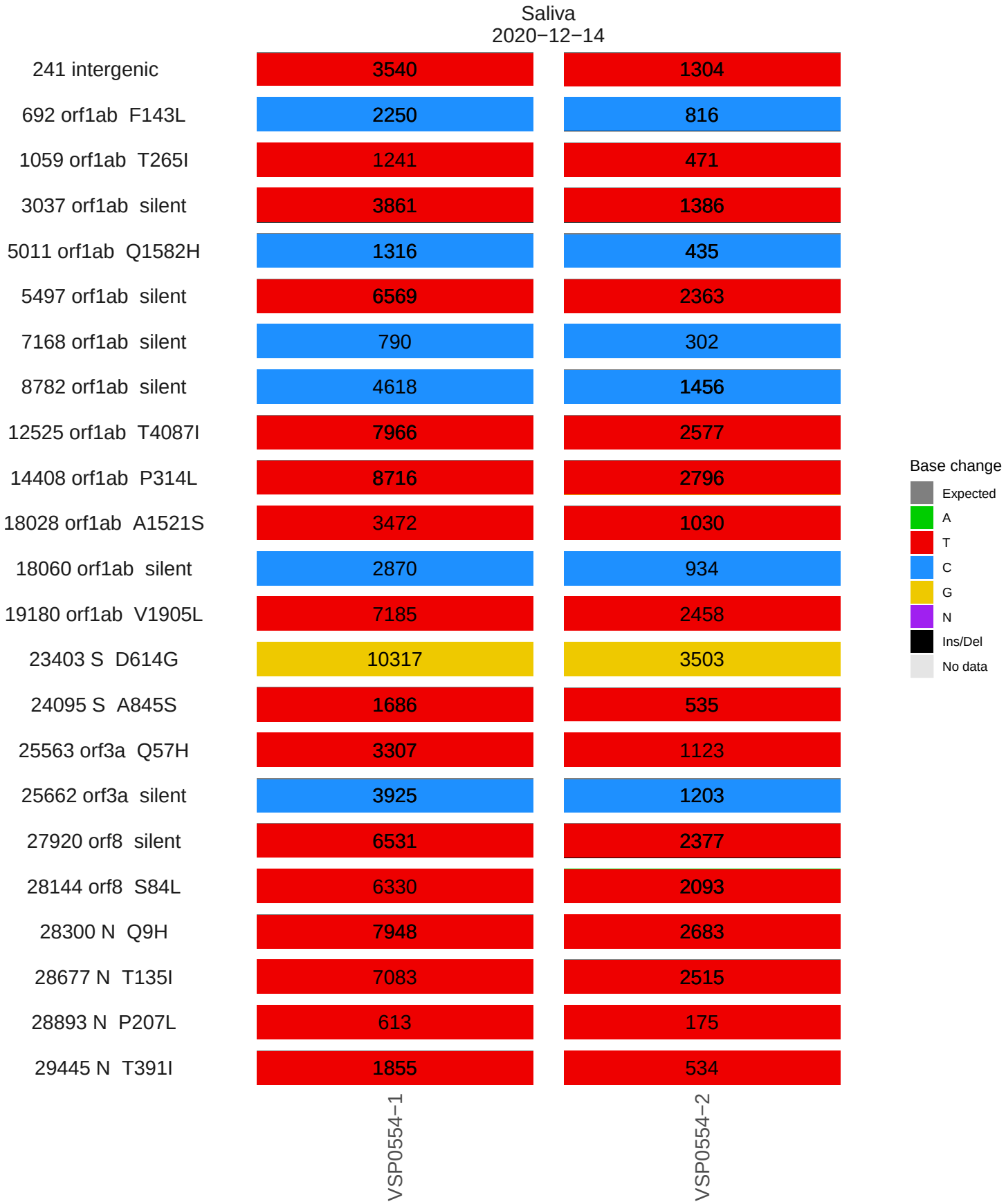
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0554	composite	NA	Saliva	2020-12-14	29.84	B.1.311	99.8%	99.8%
VSP0554-1	single experiment	NA	Saliva	2020-12-14	29.84	B.1.311	99.8%	99.8%
VSP0554-2	single experiment	NA	Saliva	2020-12-14	19.58	B.1.311	99.8%	99.7%

Variants shared across samples

The heat map below shows how variants (reference genome 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.

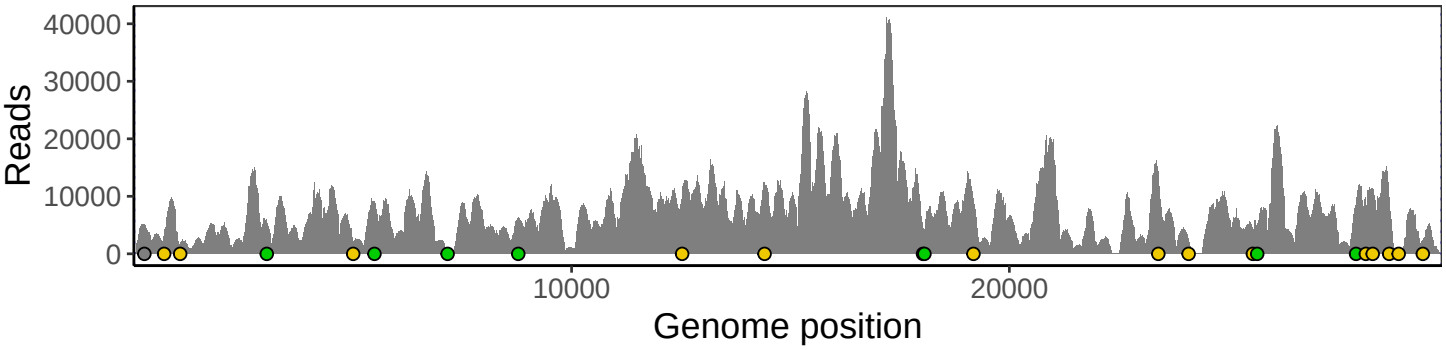



Saliva
2020−12−14
241 intergenic
692 orf1ab F143L
1059 orf1ab T265I
3037 orf1ab silent
5011 orf1ab Q1582H
5497 orf1ab silent
7168 orf1ab silent
8782 orf1ab silent
12525 orf1ab T4087I
14408 orf1ab P314L
18028 orf1ab A1521S
18060 orf1ab silent
19180 orf1ab V1905L
23403 S D614G
24095 S A845S
25563 orf3a Q57H
25662 orf3a silent
27920 orf8 silent
28144 orf8 S84L
28300 N Q9H
28677 N T135I
28893 N P207L
29445 N T391I
3540
2250
1241
3861
1316
6569
790
4618
7966
8716
3472
2870
7185
10317
1686
3307
3925
6531
6330
7948
7083
613
1855
1304
816
471
1386
435
2363
302
1456
2577
2796
1030
934
2458
3503
535
1123
1203
2377
2093
2683
2515
175
534
VSP0554−1
VSP0554−2
Base change
Expected
A
T
C
G
N
Ins/Del
No data

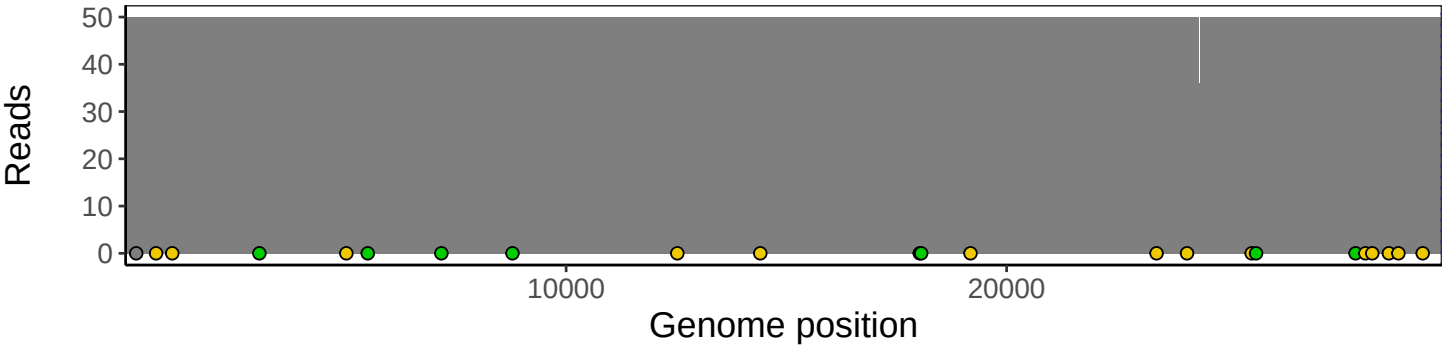
Analyses of individual experiments and composite results

VSP0554 | 2020-12-14 | Saliva | 436s | composite result

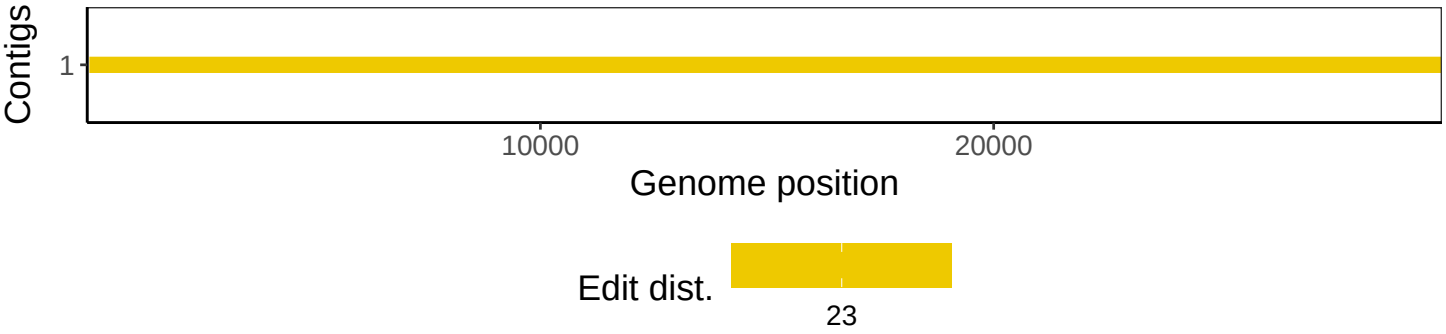
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

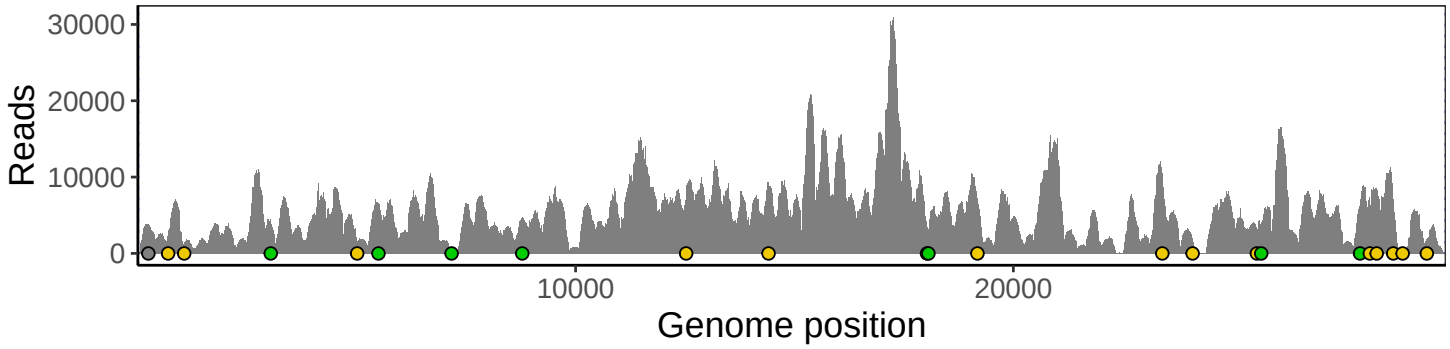


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

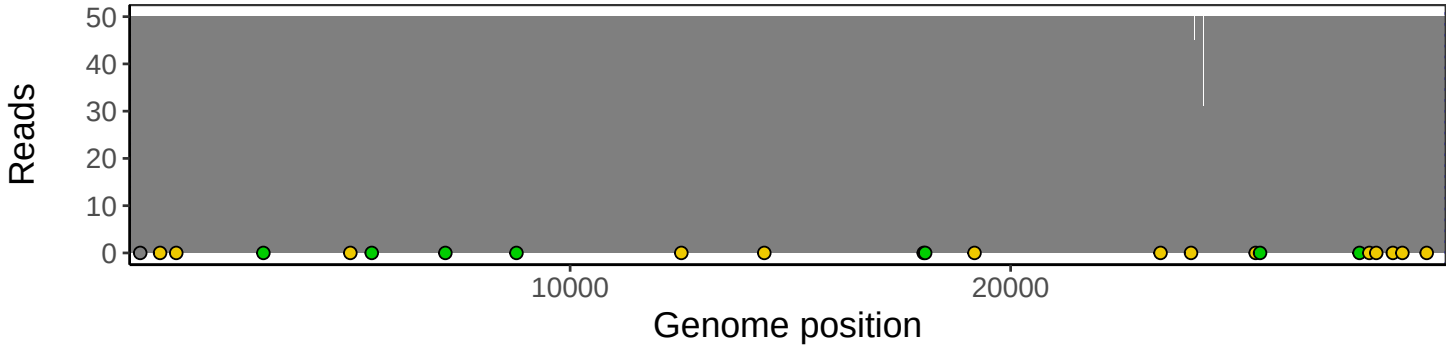


VSP0554-1 | 2020-12-14 | Saliva | 436s | genomes | single experiment

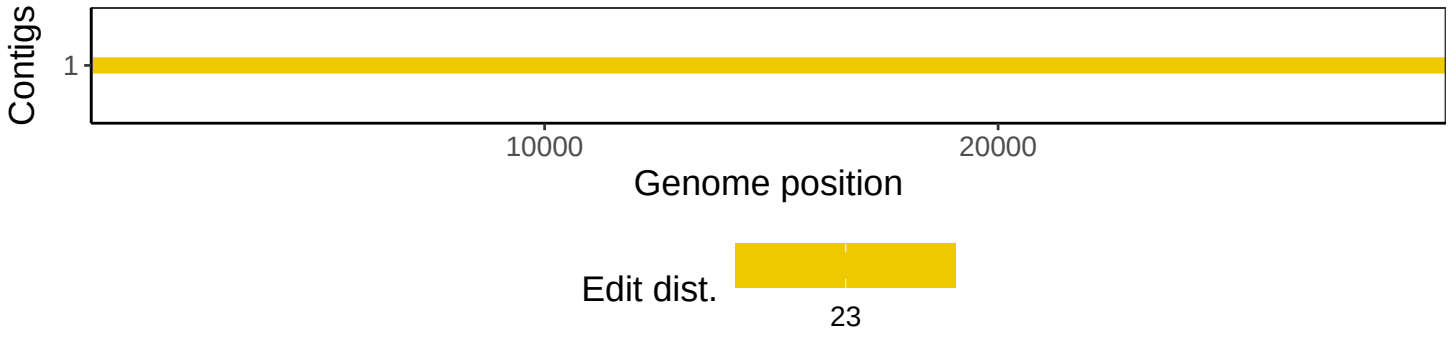
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

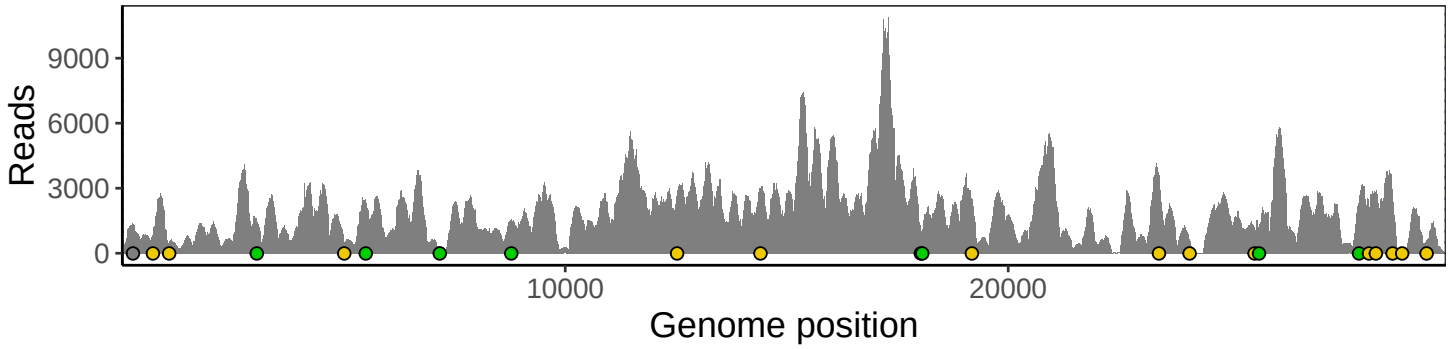


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

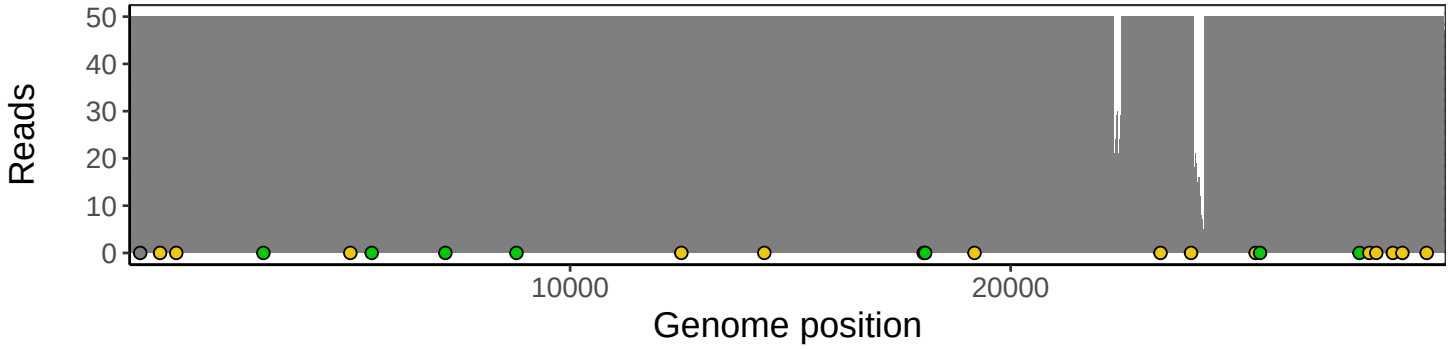


VSP0554-2 | 2020-12-14 | Saliva | 436s | genomes | single experiment

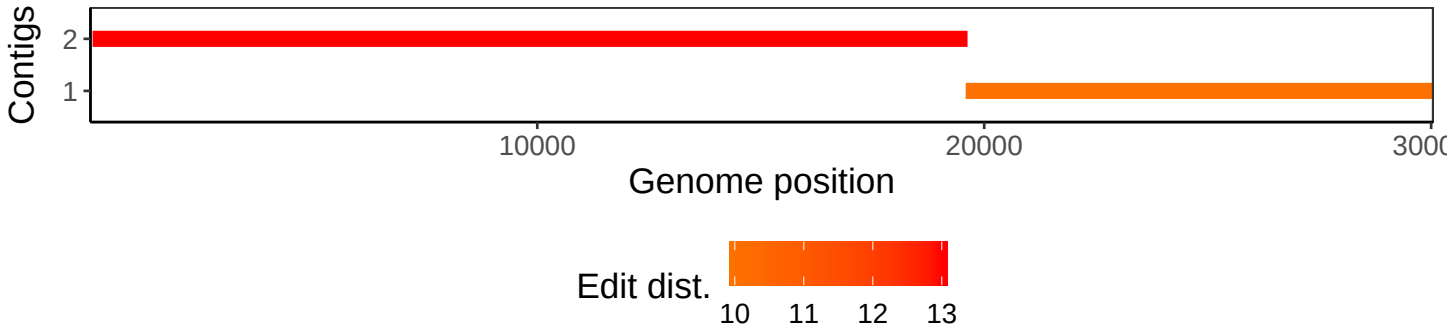
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.3.3
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	3.0.0
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