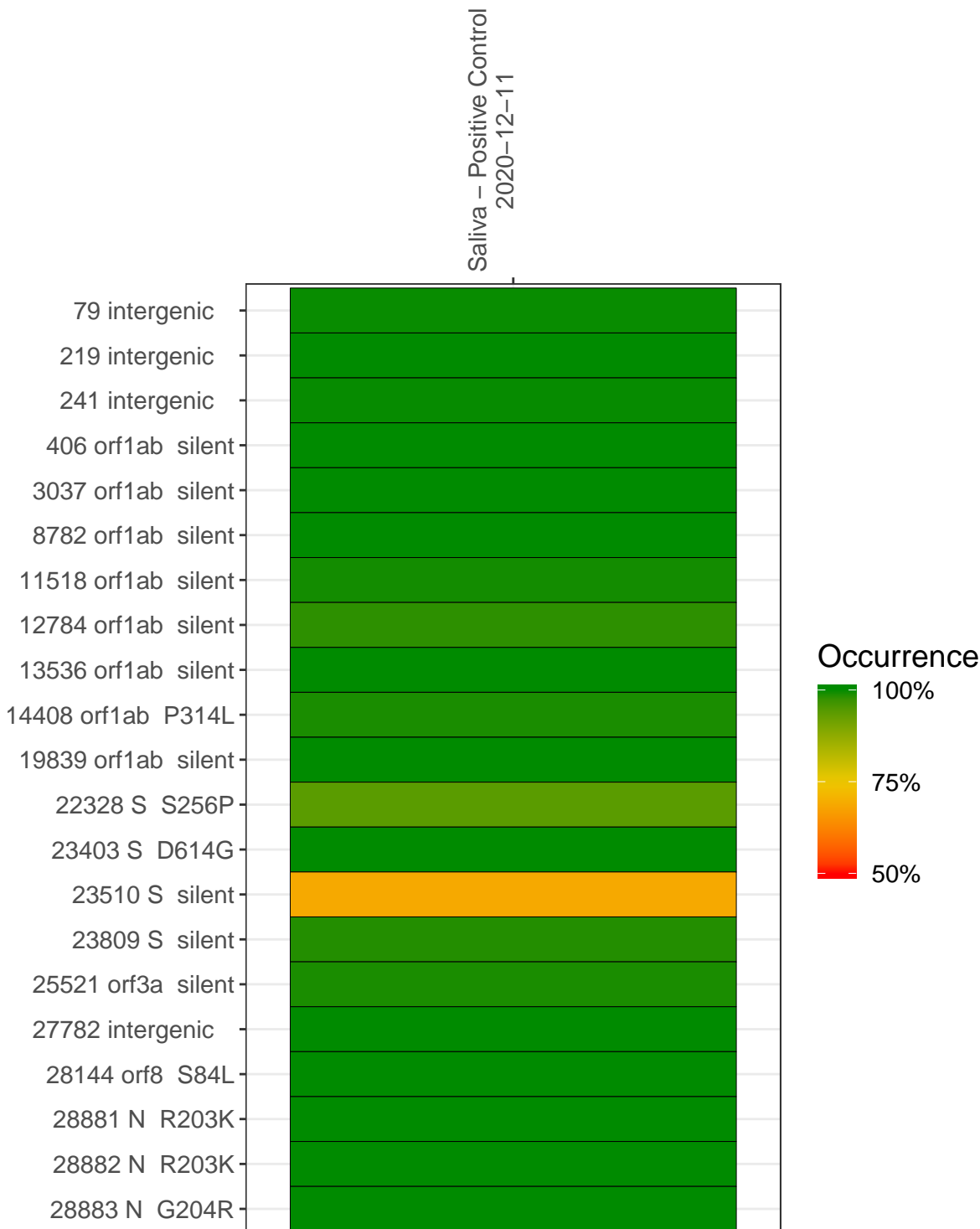# COVID-19 subject 10-1017

2021-01-31

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.
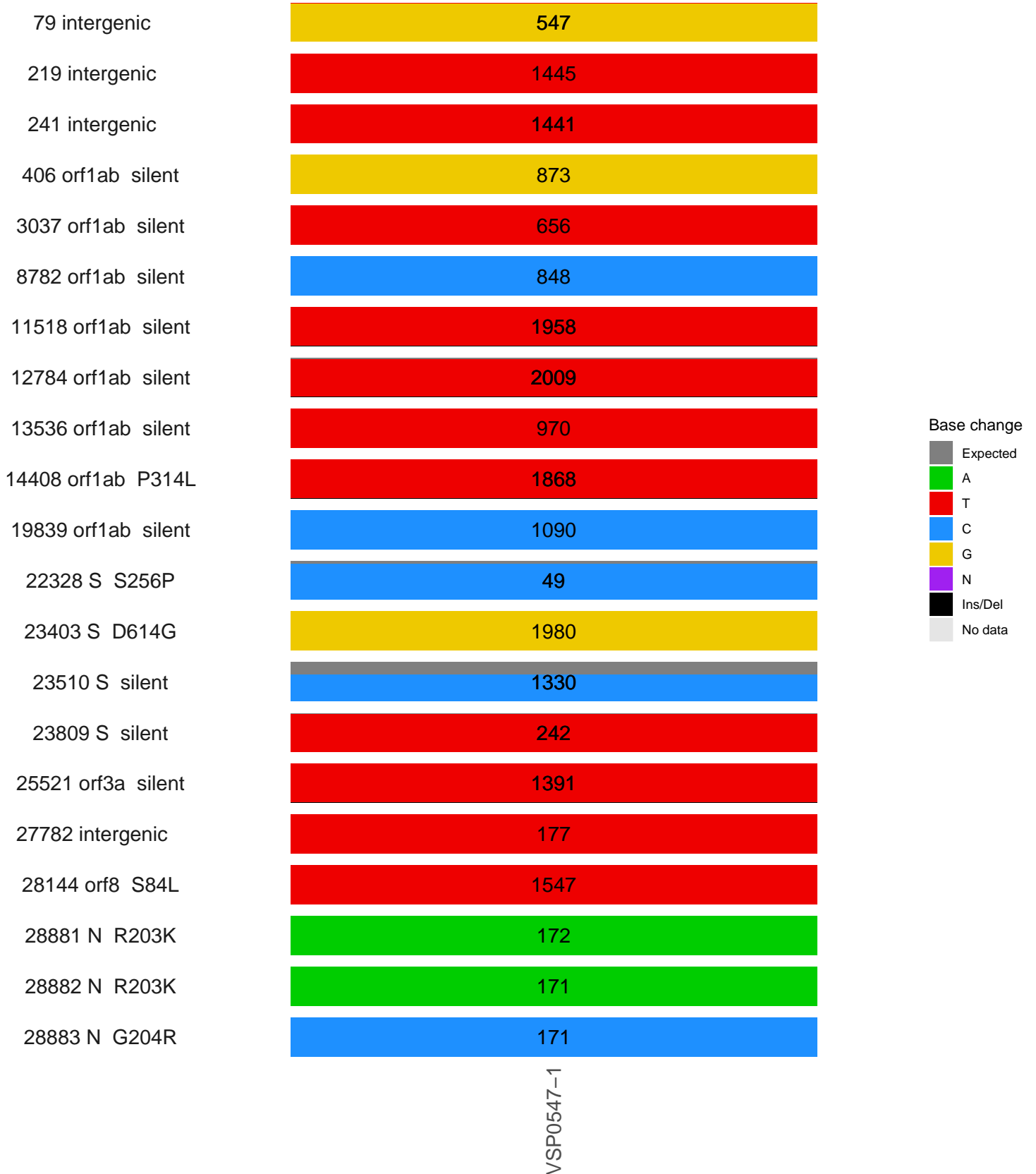
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0547-1	single experiment	NA	Saliva - Positive Control	2020-12-11	29.85	B.1.1	99.9%	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 – Positive Control
2020−12−11

Position	Value
79 intergenic	547
219 intergenic	1445
241 intergenic	1441
406 orf1ab silent	873
3037 orf1ab silent	656
8782 orf1ab silent	848
11518 orf1ab silent	1958
12784 orf1ab silent	2009
13536 orf1ab silent	970
14408 orf1ab P314L	1868
19839 orf1ab silent	1090
22328 S S256P	49
23403 S D614G	1980
23510 S silent	1330
23809 S silent	242
25521 orf3a silent	1391
27782 intergenic	177
28144 orf8 S84L	1547
28881 N R203K	172
28882 N R203K	171
28883 N G204R	171

Base change
- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

VSP0547−1

3

Analyses of individual experiments and composite results

VSP0547-1 | 2020-12-11 | Saliva - Positive Control | 10-1017 | 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.
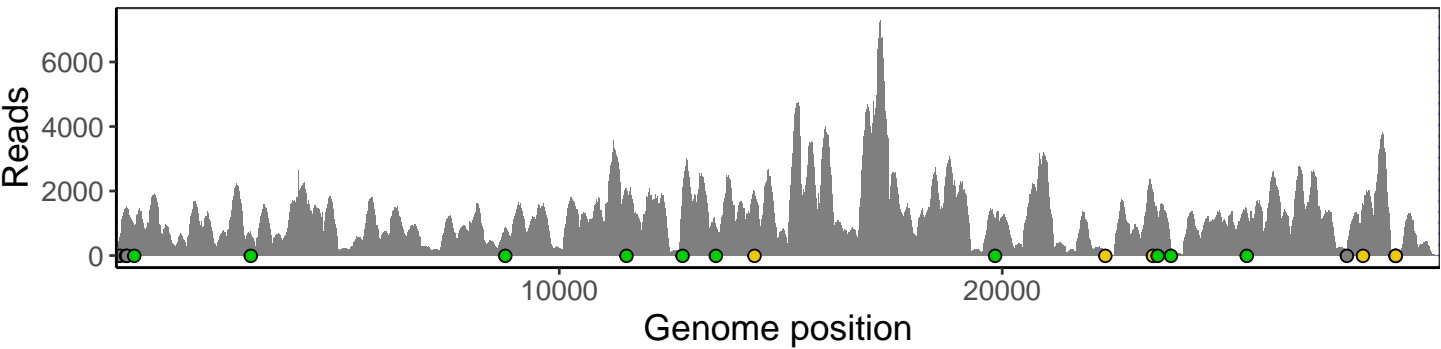
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
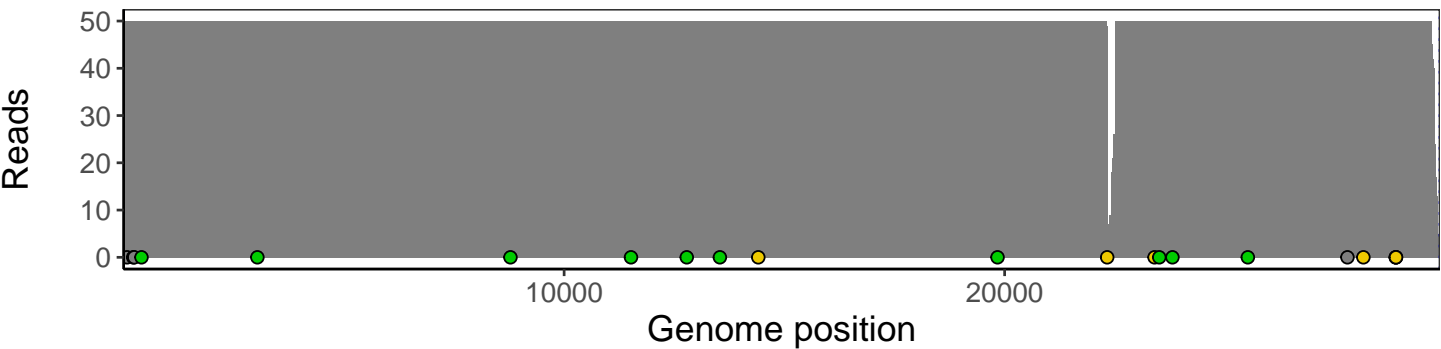
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
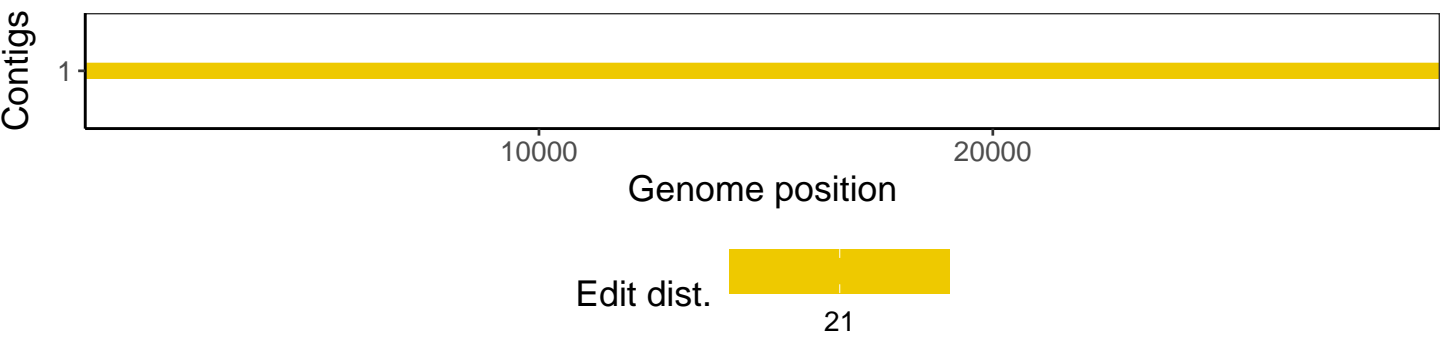
Software environment

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	2.1.7
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