COVID-19 subject UPHS-0667

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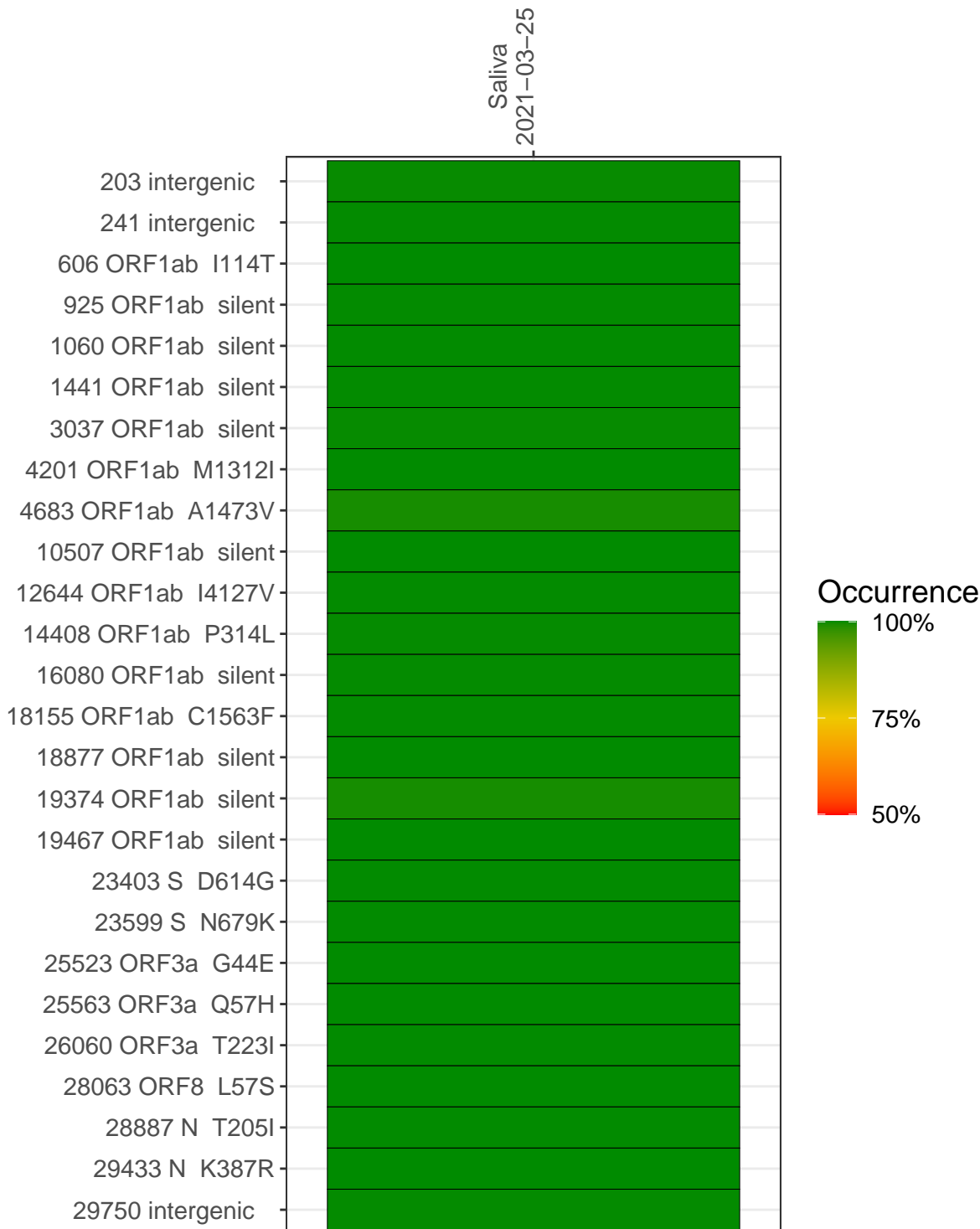 sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with > 90% sequence coverage.
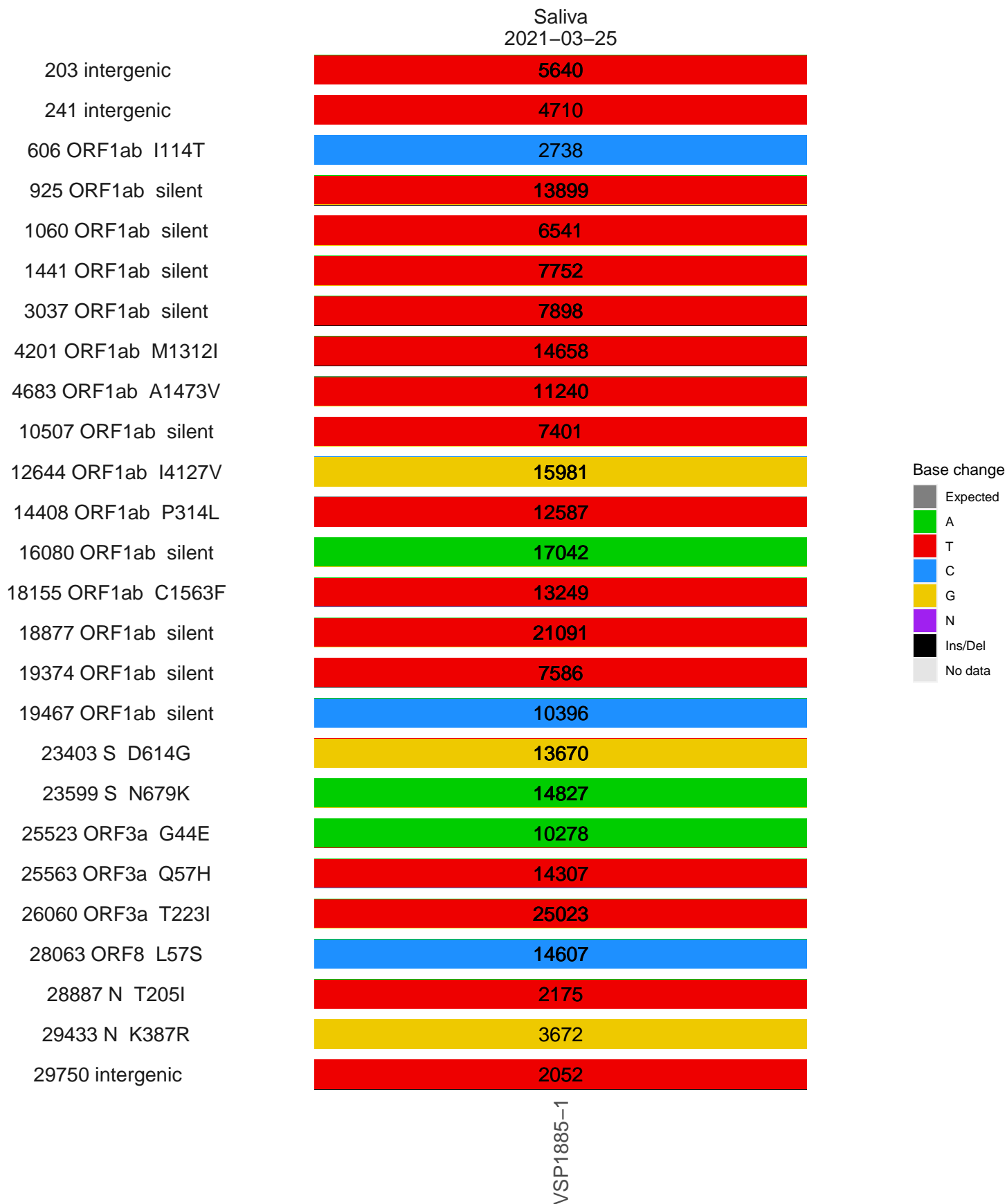
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1885-1	single experiment	NA	Saliva	2021-03-25	29.89	B.1.111	99.8%	99.8%

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
2021−03−25

Position	Value
203 intergenic	5640
241 intergenic	4710
606 ORF1ab I114T	2738
925 ORF1ab silent	13899
1060 ORF1ab silent	6541
1441 ORF1ab silent	7752
3037 ORF1ab silent	7898
4201 ORF1ab M1312I	14658
4683 ORF1ab A1473V	11240
10507 ORF1ab silent	7401
12644 ORF1ab I4127V	15981
14408 ORF1ab P314L	12587
16080 ORF1ab silent	17042
18155 ORF1ab C1563F	13249
18877 ORF1ab silent	21091
19374 ORF1ab silent	7586
19467 ORF1ab silent	10396
23403 S D614G	13670
23599 S N679K	14827
25523 ORF3a G44E	10278
25563 ORF3a Q57H	14307
26060 ORF3a T223I	25023
28063 ORF8 L57S	14607
28887 N T205I	2175
29433 N K387R	3672
29750 intergenic	2052

VSP1885−1

Base change

(gray)	Expected
(green)	A
(red)	T
(blue)	C
(yellow)	G
(purple)	N
(black)	Ins/Del
(light gray)	No data

3

Analyses of individual experiments and composite results

**VSP1885-1 | 2021-03-25 | Saliva | UPHS-0667 | 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.
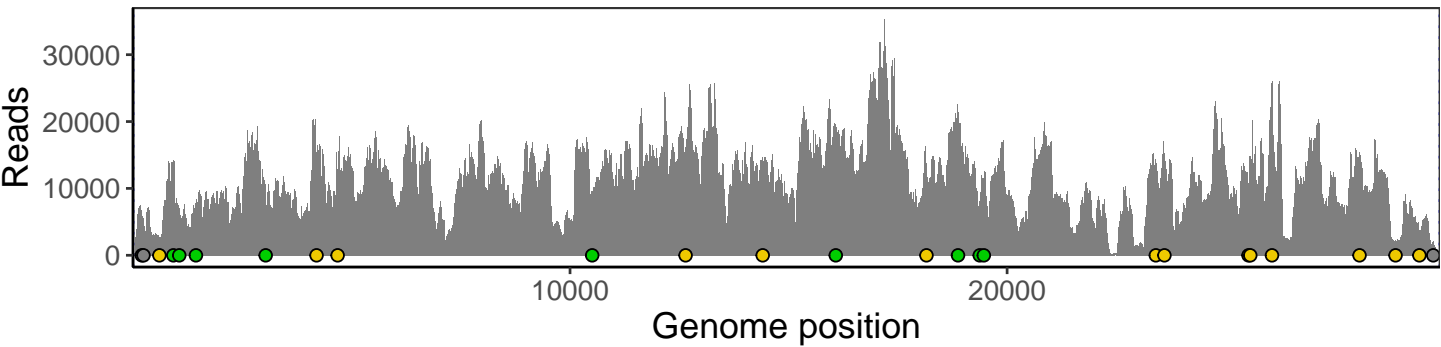
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
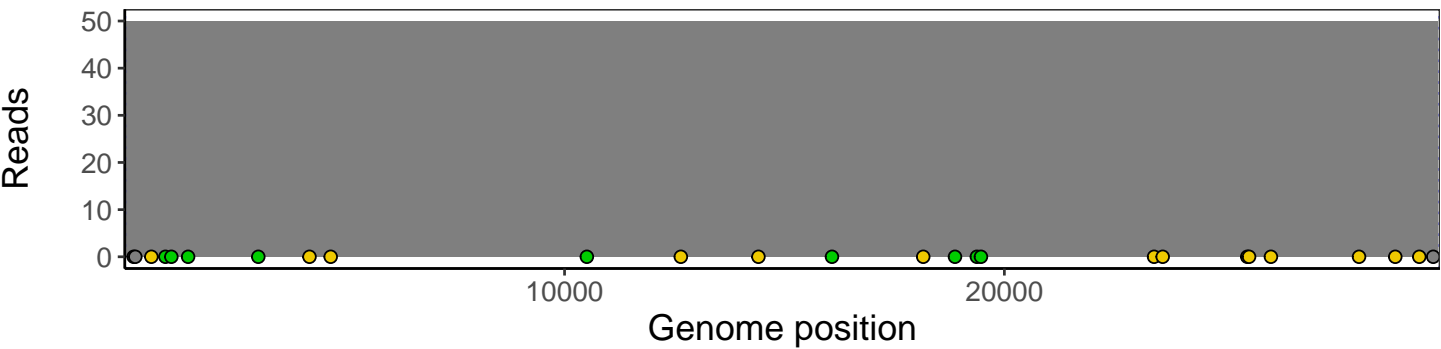
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
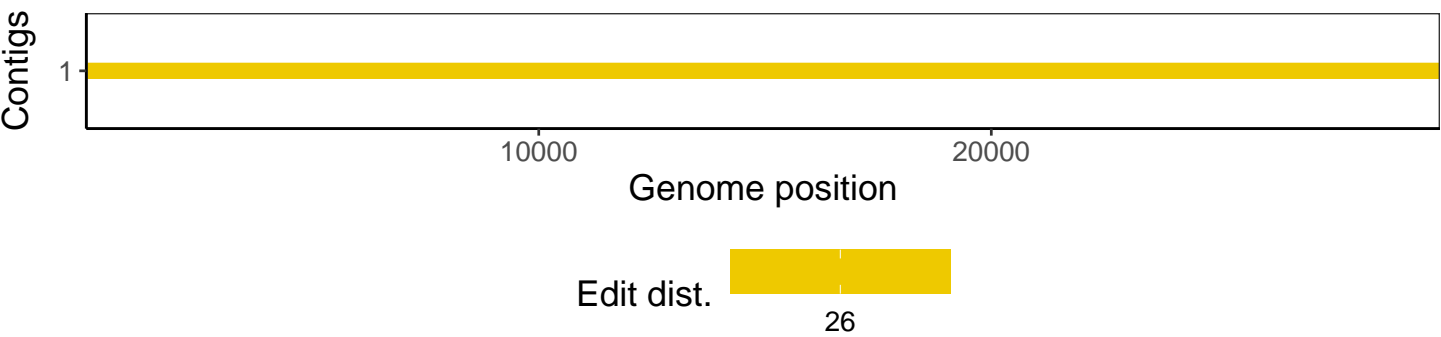
Software environment

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