

COVID-19 subject HUP Q-0019

2021-03-29

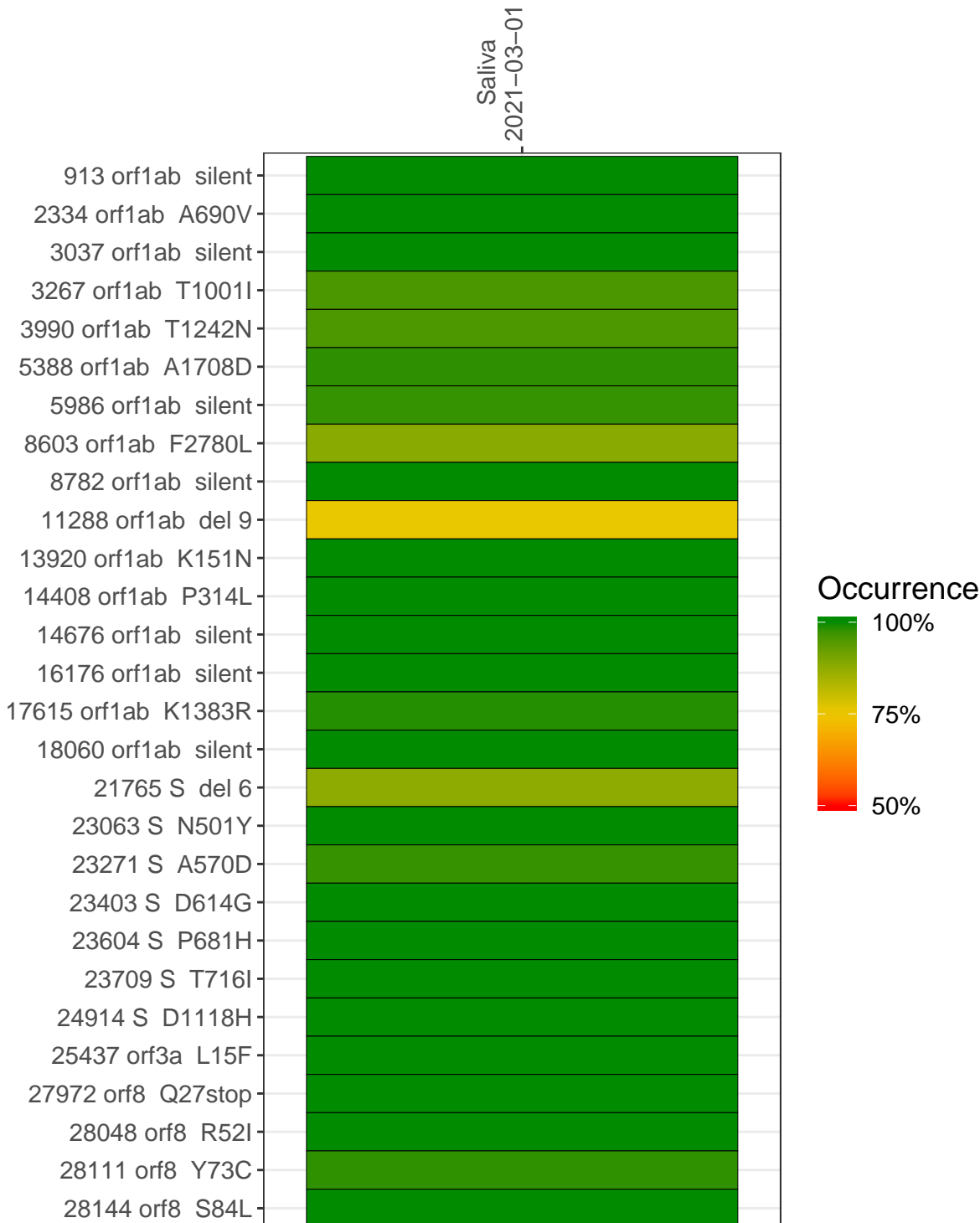
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP0883-1	single experiment	NA	Saliva	2021-03-01	10.39	B.1.1.7	99.0%	97.0%

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

913 orf1ab silent	23
2334 orf1ab A690V	39
3037 orf1ab silent	39
3267 orf1ab T1001I	22
3990 orf1ab T1242N	21
5388 orf1ab A1708D	51
5986 orf1ab silent	40
8603 orf1ab F2780L	51
8782 orf1ab silent	24
11288 orf1ab del 9	31
13920 orf1ab K151N	72
14408 orf1ab P314L	75
14676 orf1ab silent	16
16176 orf1ab silent	73
17615 orf1ab K1383R	75
18060 orf1ab silent	53
21765 S del 6	49
23063 S N501Y	62
23271 S A570D	37
23403 S D614G	30
23604 S P681H	81
23709 S T716I	58
24914 S D1118H	136
25437 orf3a L15F	68
27972 orf8 Q27stop	95
28048 orf8 R52I	93
28111 orf8 Y73C	93
28144 orf8 S84L	50

Base change

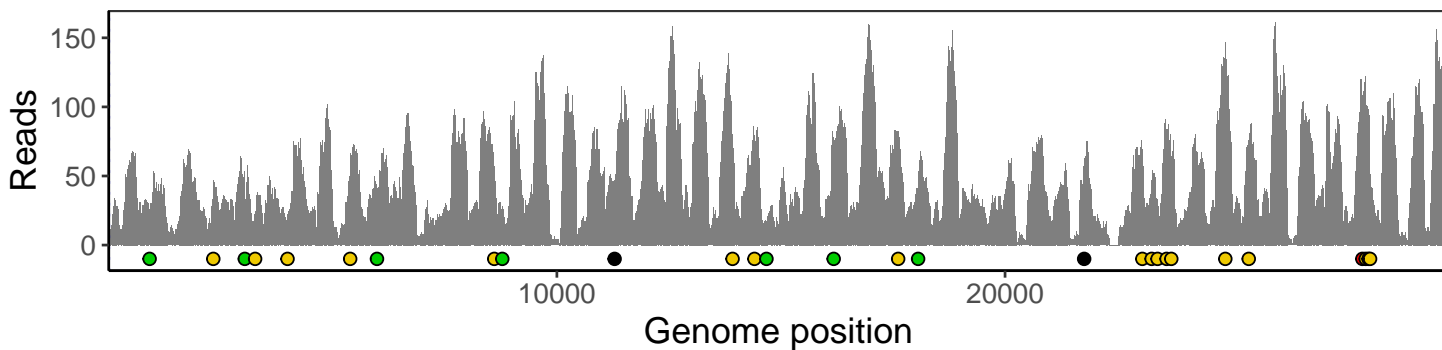
Expected
A
T
C
G
N
Ins/Del
No data

VSP0883-1

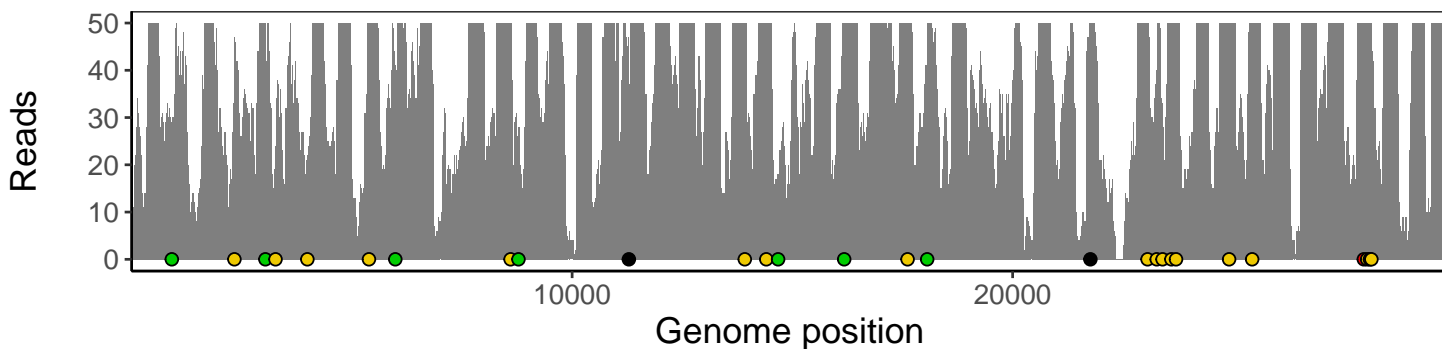
Analyses of individual experiments and composite results

VSP0883-1 | 2021-03-01 | Saliva | HUP Q-0019 | genomes | single experiment

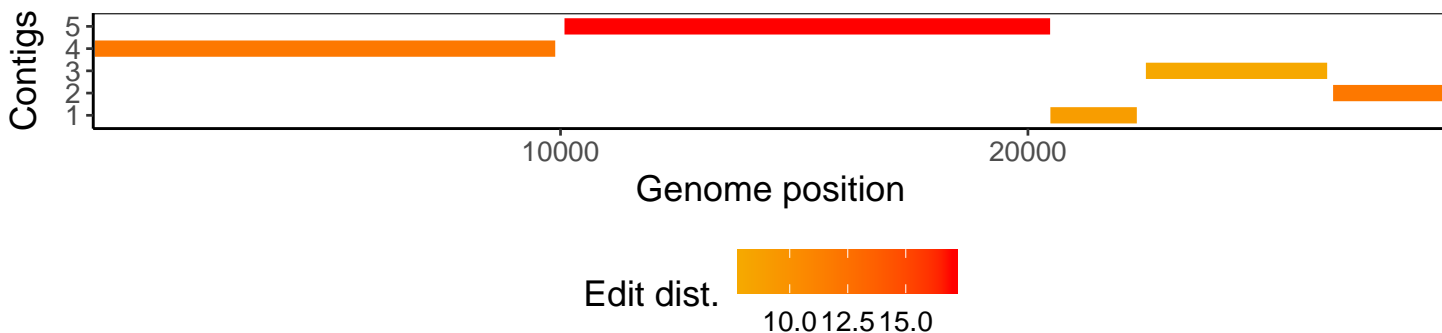
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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 htlib 1.10
bcftools	1.10.2-34-g1a12af0-dirty Using htlib 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