COVID-19 subject 445

2021-06-23

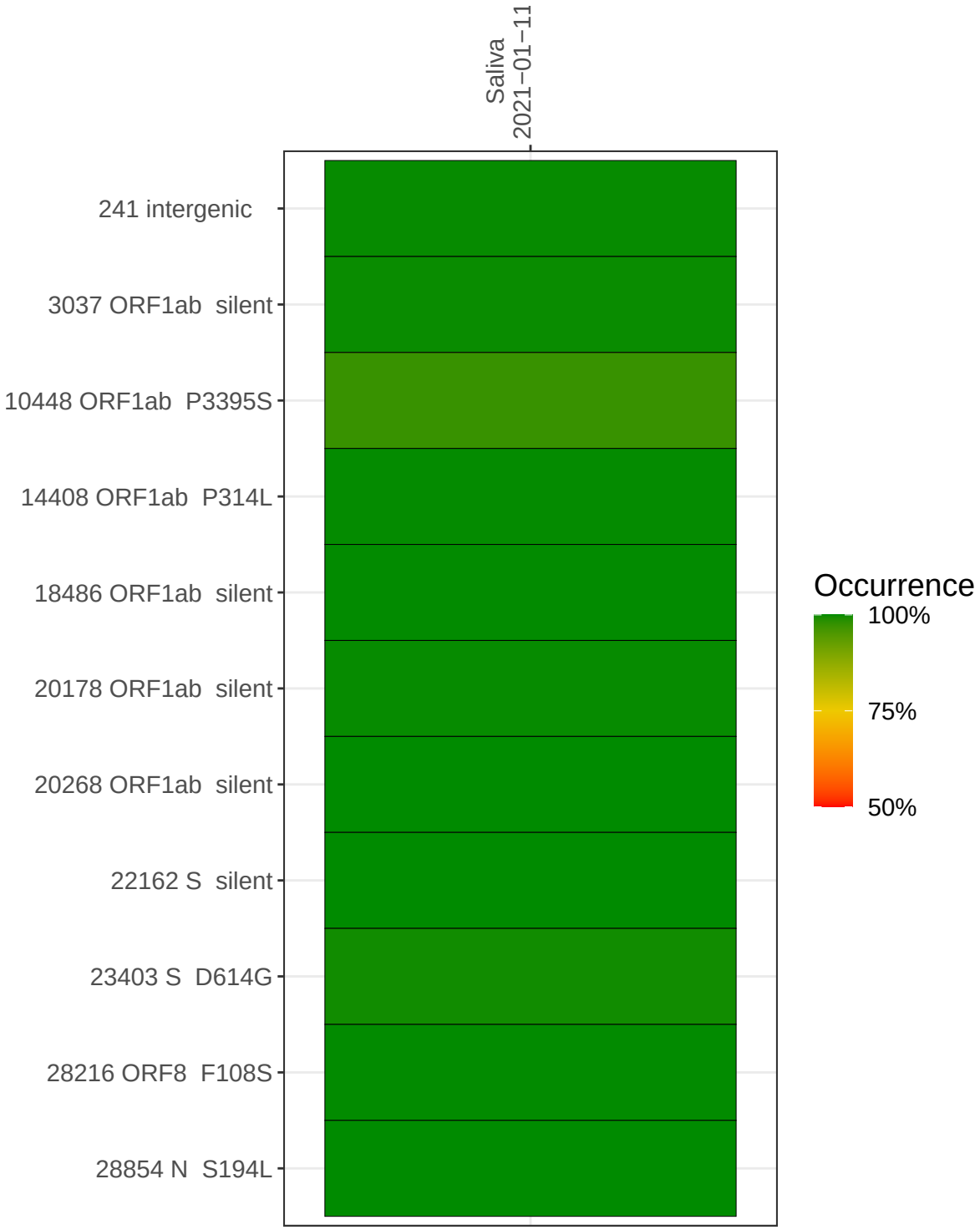
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

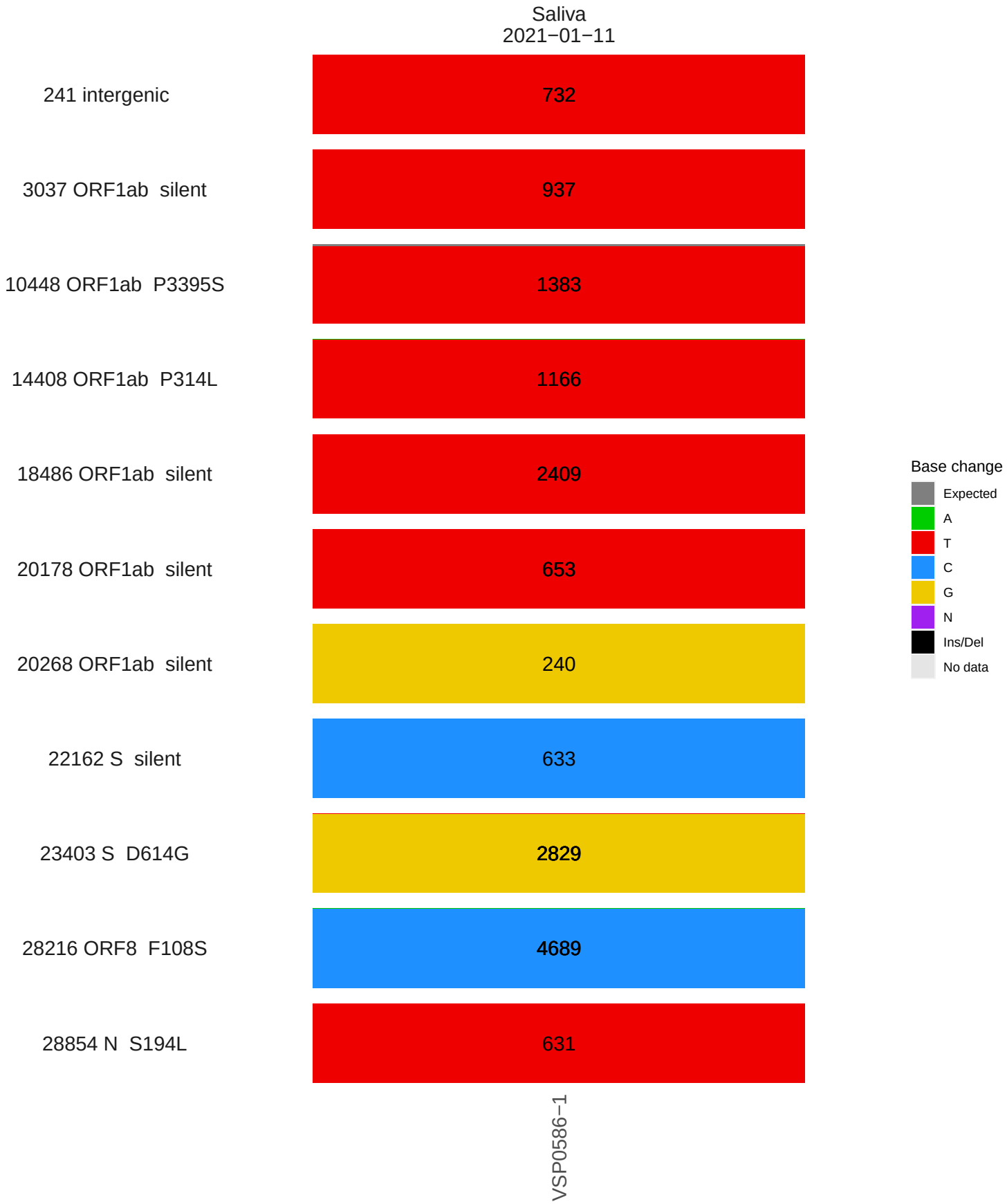
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0586-1	single experiment	NA	Saliva	2021-01-11	29.90	B.1.240	99.9%	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.

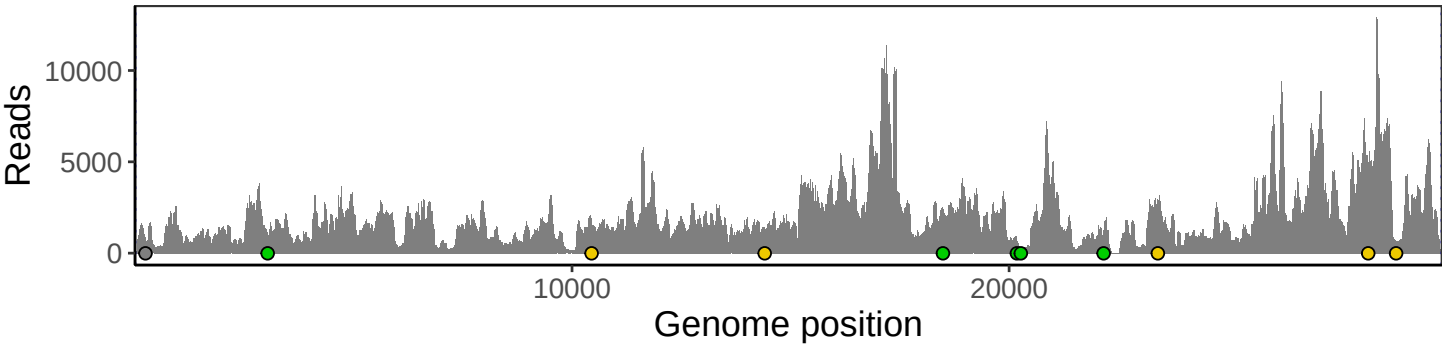



Saliva
2021−01−11
241 intergenic
732
3037 ORF1ab silent
937
10448 ORF1ab P3395S
1383
14408 ORF1ab P314L
1166
18486 ORF1ab silent
2409
20178 ORF1ab silent
653
20268 ORF1ab silent
240
22162 S silent
633
23403 S D614G
2829
28216 ORF8 F108S
4689
28854 N S194L
631
VSP0586−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

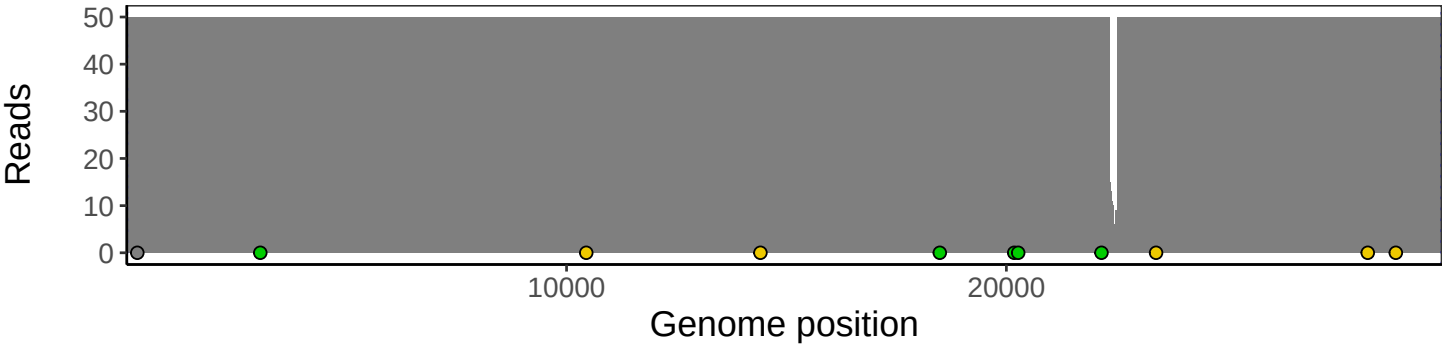
Analyses of individual experiments and composite results

VSP0586-1 | 2021-01-11 | Saliva | 445s | genomes | single experiment

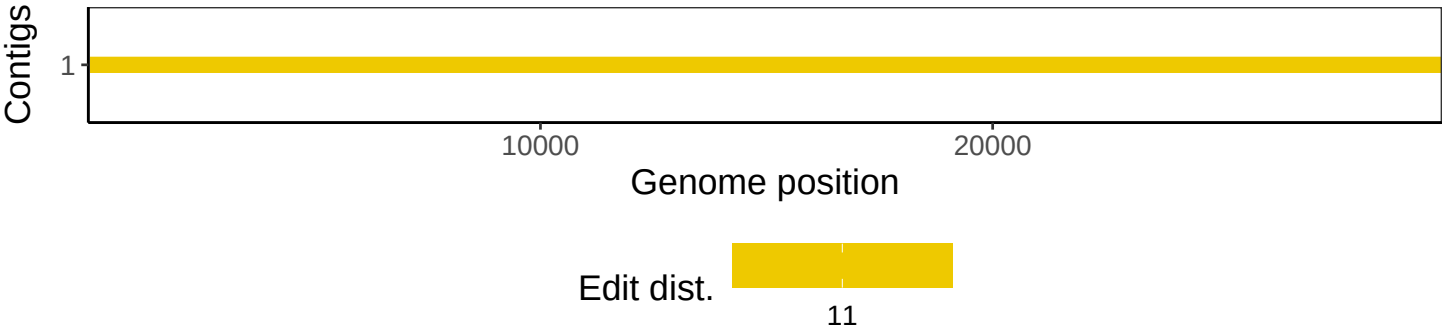
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	1.6.1
optparse	1.6.6
forcats	0.5.1
stringr	1.4.0
dplyr	1.0.7
purrr	0.3.4
readr	1.4.0
tidyr	1.1.3
tibble	3.1.2
ggplot2	3.3.4
tidyverse	1.3.1
ShortRead	1.36.1
GenomicAlignments	1.14.2
SummarizedExperiment	1.8.1
DelayedArray	0.4.1
matrixStats	0.59.0
Biobase	2.38.0
Rsamtools	1.30.0
GenomicRanges	1.30.3
GenomeInfoDb	1.14.0
Biostrings	2.46.0
XVector	0.18.0
IRanges	2.12.0
S4Vectors	0.16.0
BiocParallel	1.12.0
BiocGenerics	0.24.0