COVID-19 subject UPHS-1360

2021-05-21

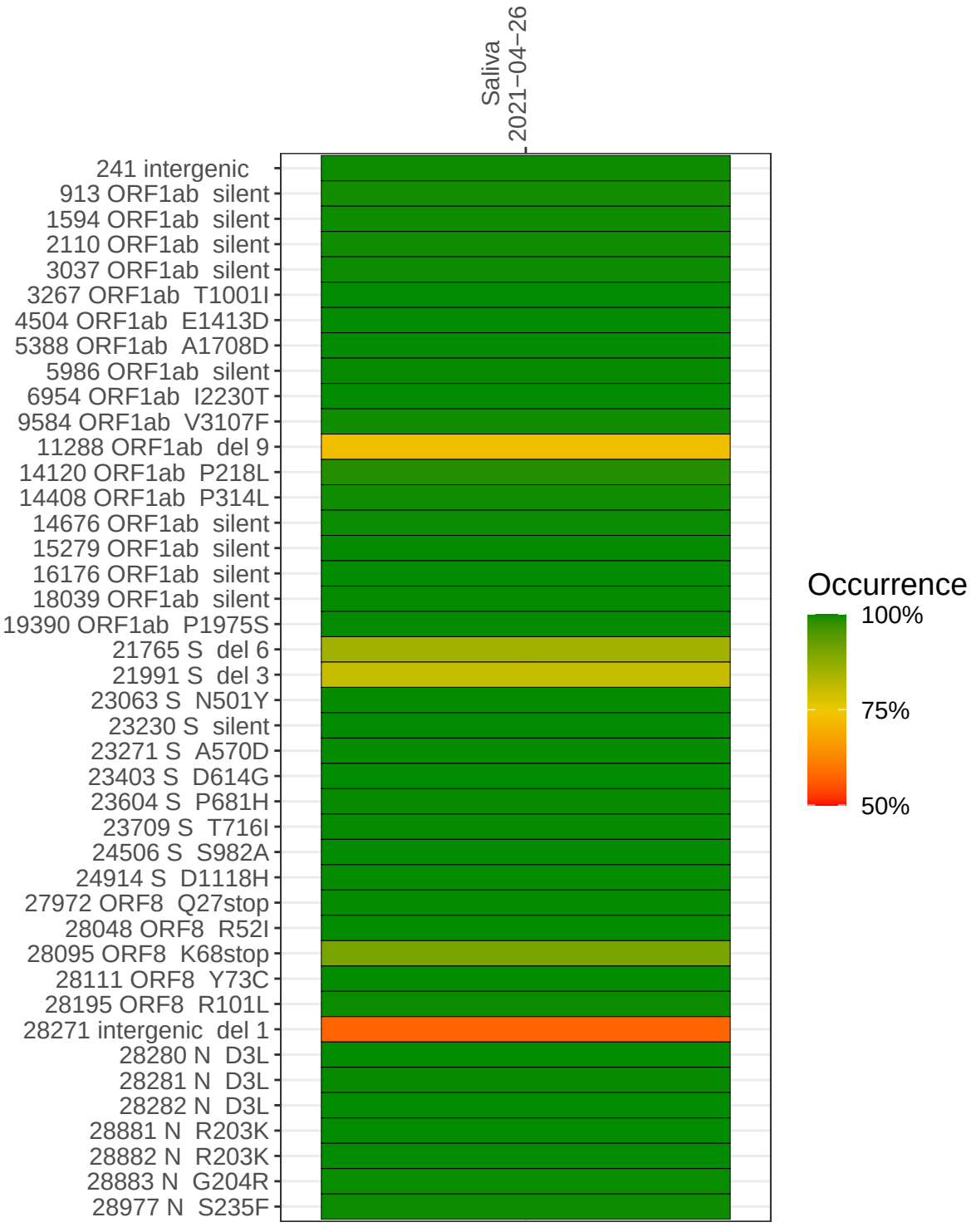
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

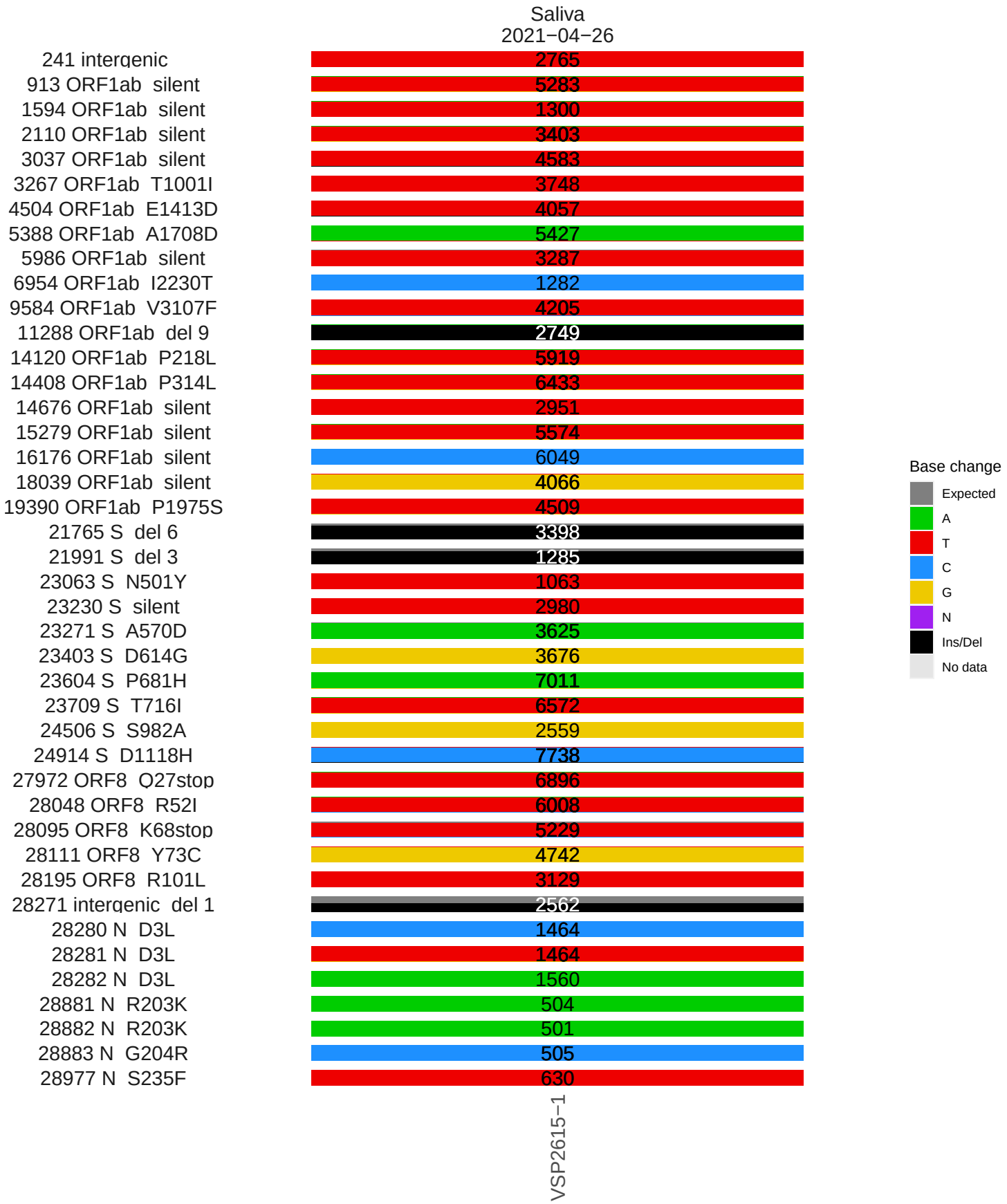
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP2615-1	single experiment	NA	Saliva	2021-04-26	29.82	B.1.1.7	99.8%	99.7%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/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.

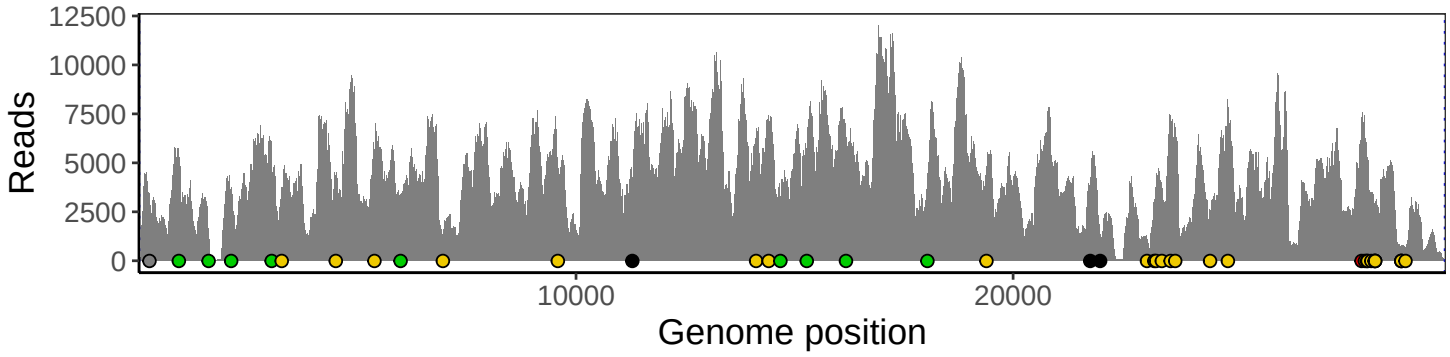



Saliva
2021−04−26
241 intergenic 2765
913 ORF1ab silent 5283
1594 ORF1ab silent 1300
2110 ORF1ab silent 3403
3037 ORF1ab silent 4583
3267 ORF1ab T1001I 3748
4504 ORF1ab E1413D 4057
5388 ORF1ab A1708D 5427
5986 ORF1ab silent 3287
6954 ORF1ab I2230T 1282
9584 ORF1ab V3107F 4205
11288 ORF1ab del 9 2749
14120 ORF1ab P218L 5919
14408 ORF1ab P314L 6433
14676 ORF1ab silent 2951
15279 ORF1ab silent 5574
16176 ORF1ab silent 6049
18039 ORF1ab silent 4066
19390 ORF1ab P1975S 4509
21765 S del 6 3398
21991 S del 3 1285
23063 S N501Y 1063
23230 S silent 2980
23271 S A570D 3625
23403 S D614G 3676
23604 S P681H 7011
23709 S T716I 6572
24506 S S982A 2559
24914 S D1118H 7738
27972 ORF8 Q27stop 6896
28048 ORF8 R52I 6008
28095 ORF8 K68stop 5229
28111 ORF8 Y73C 4742
28195 ORF8 R101L 3129
28271 intergenic del 1 2562
28280 N D3L 1464
28281 N D3L 1464
28282 N D3L 1560
28881 N R203K 504
28882 N R203K 501
28883 N G204R 505
28977 N S235F 630
VSP2615−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

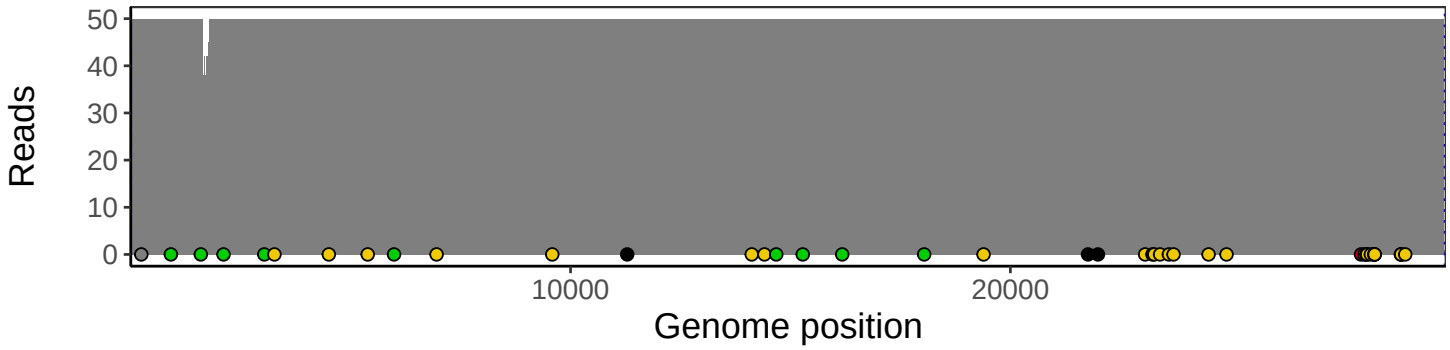
Analyses of individual experiments and composite results

VSP2615-1 | 2021-04-26 | Saliva | UPHS-1360 | genomes | single experiment

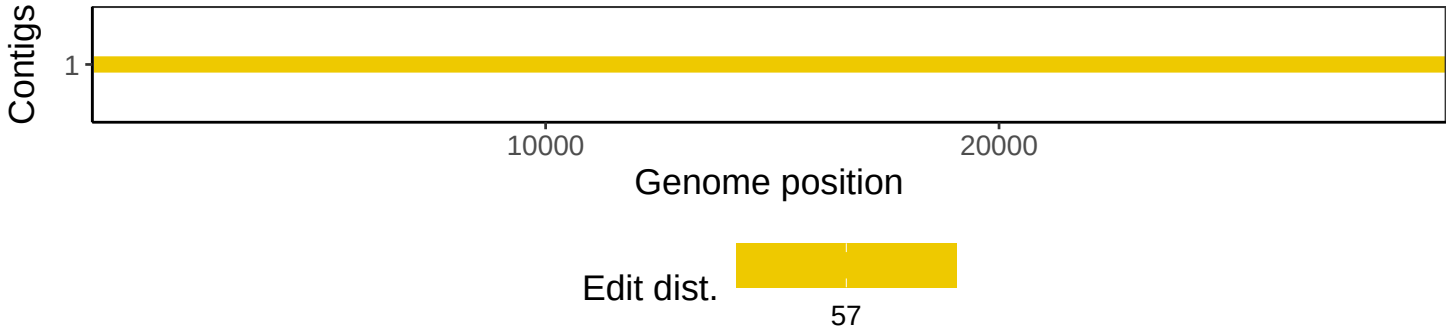
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.3.8
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