

COVID-19 subject 2757

2021-01-08

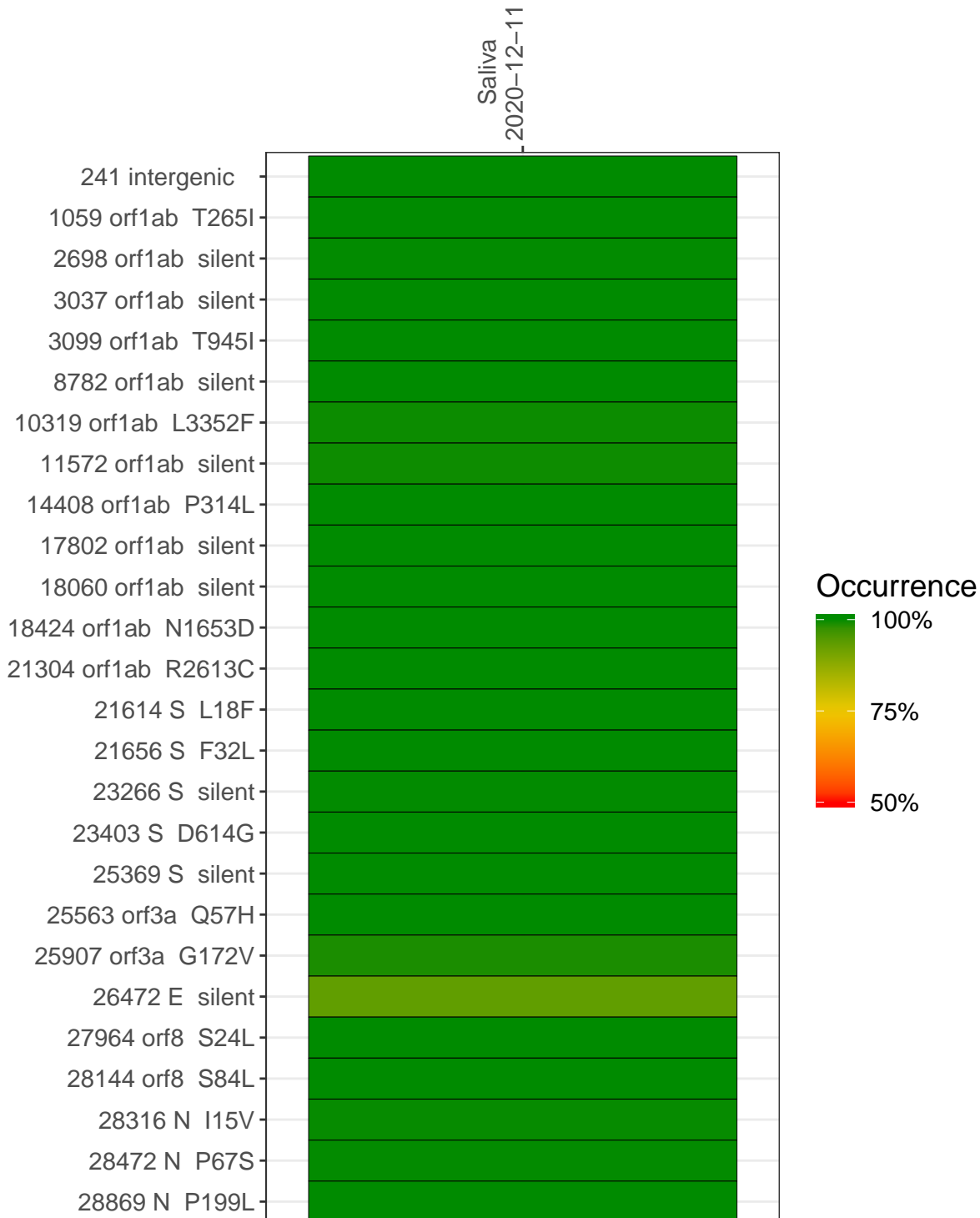
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. The code base for this analysis can be found ([here](#)).

Table 1. Sample summary.

Experiment	Type	Input genomes	Sample type	Sample date	Largest contig (KD)	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0538-1	single experiment	NA	Saliva	2020-12-11	22.40	100.0%	98.3%

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 for 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
2020-12-11

241 intergenic	859
1059 orf1ab T265I	525
2698 orf1ab silent	1821
3037 orf1ab silent	271
3099 orf1ab T945I	227
8782 orf1ab silent	267
10319 orf1ab L3352F	1434
11572 orf1ab silent	302
14408 orf1ab P314L	708
17802 orf1ab silent	841
18060 orf1ab silent	223
18424 orf1ab N1653D	1548
21304 orf1ab R2613C	25
21614 S L18F	44
21656 S F32L	40
23266 S silent	1757
23403 S D614G	2123
25369 S silent	889
25563 orf3a Q57H	1096
25907 orf3a G172V	577
26472 E silent	636
27964 orf8 S24L	600
28144 orf8 S84L	951
28316 N I15V	1413
28472 N P67S	1426
28869 N P199L	235

Base change

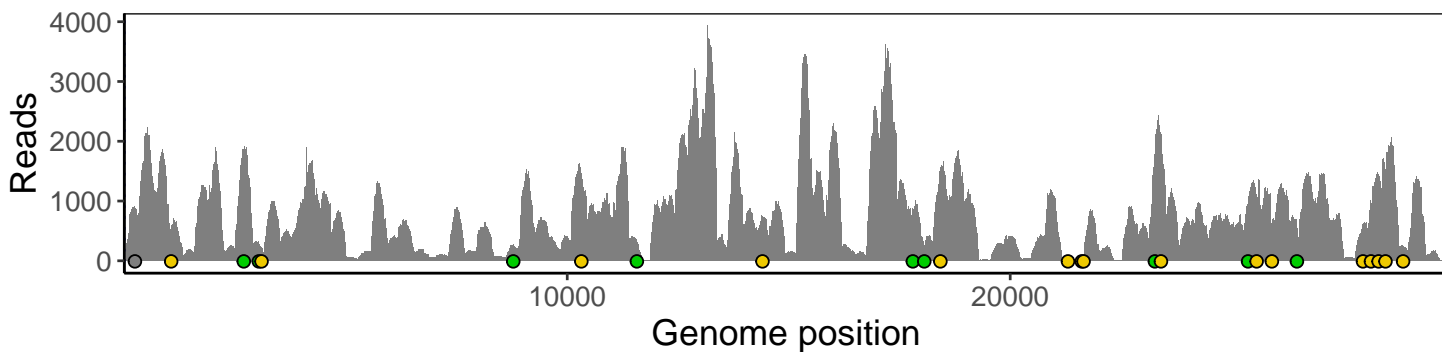
Expected	
A	
T	
C	
G	
N	
Ins/Del	
No data	

VSP0538-1

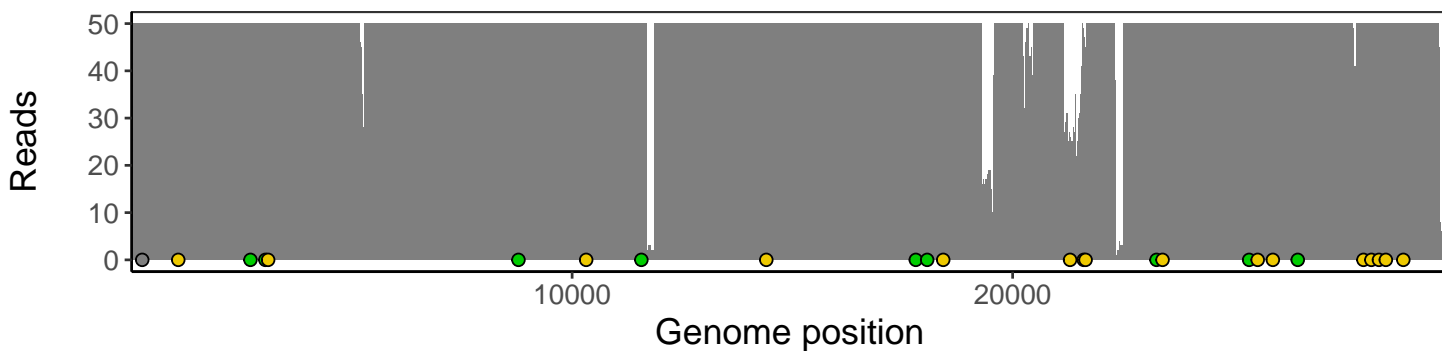
Analyses of individual experiments and composite results.

VSP0538-1 | 2020-12-11 | Saliva | 2757 | 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.

