COVID-19 subject England_EPI_ISL_747519

2021-01-08

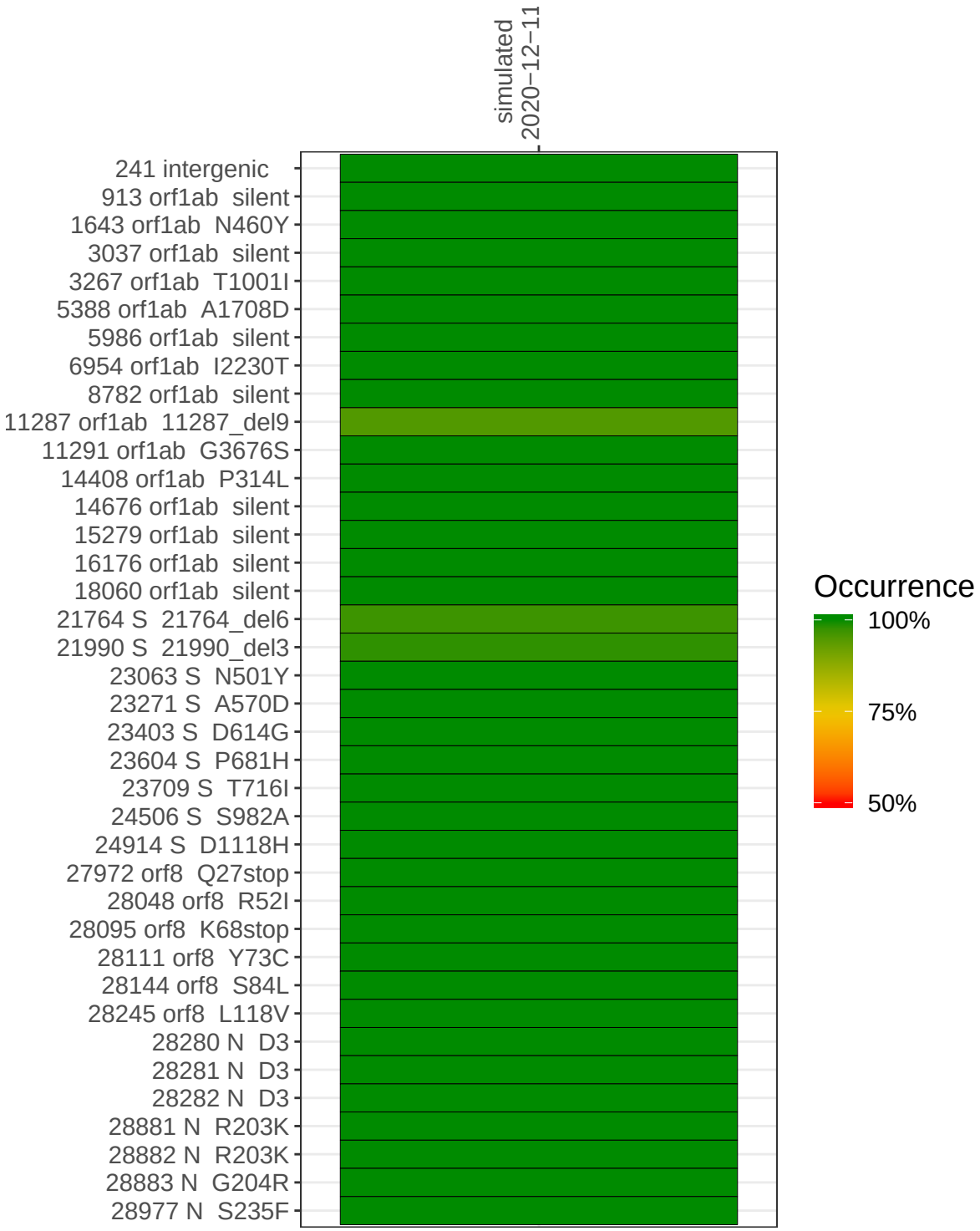
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. The code base for this analysis can be found (here).

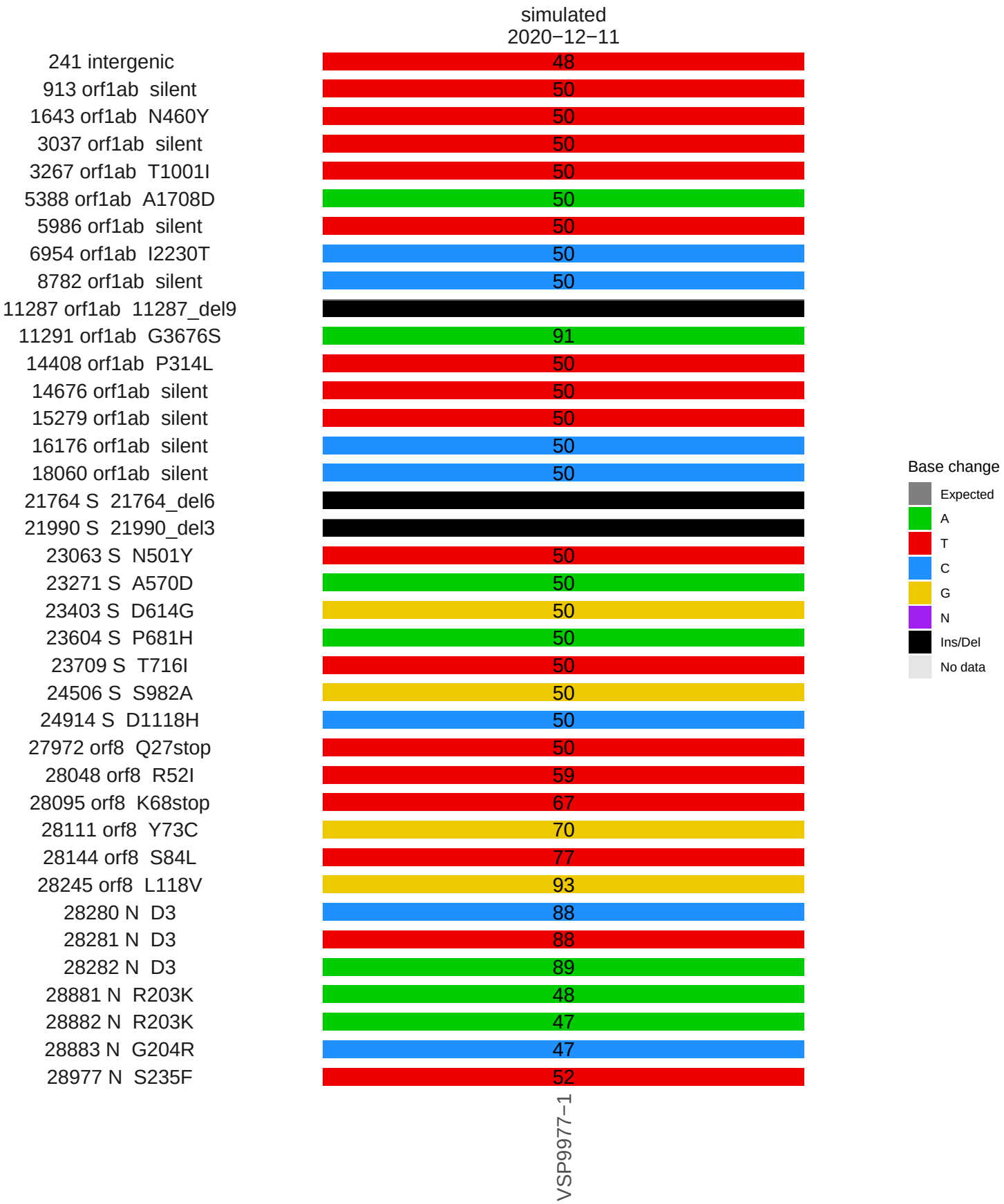
Table 1. Sample summary.

Experiment	Type	Input genomes	Sample type	Sample date	Largest contig (KD)	Reference read coverage	Reference read coverage (>= 5 reads)
VSP9977-1	single experiment	NA	simulated	2020-12-11	29.84	100.0%	99.8%

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.

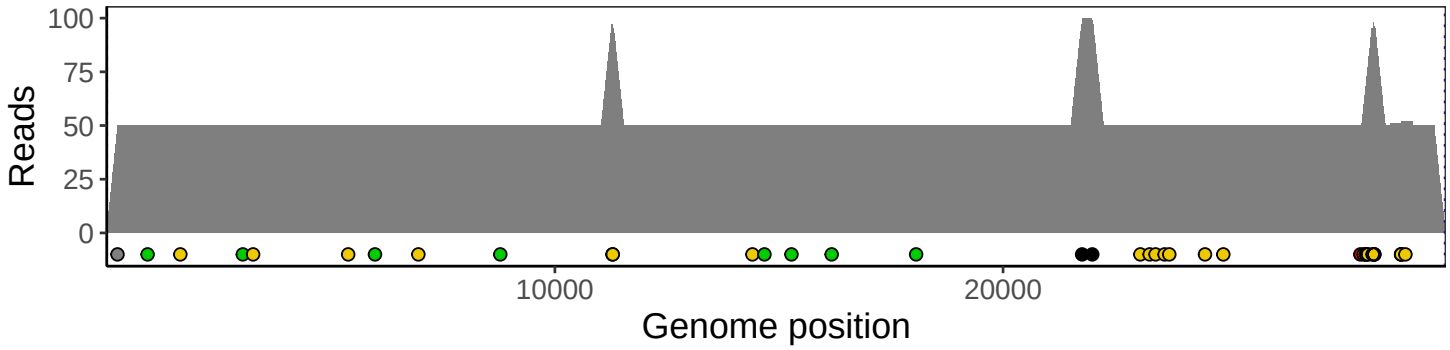



simulated
2020−12−11
241 intergenic
913 orf1ab silent
1643 orf1ab N460Y
3037 orf1ab silent
3267 orf1ab T1001I
5388 orf1ab A1708D
5986 orf1ab silent
6954 orf1ab I2230T
8782 orf1ab silent
11287 orf1ab 11287_del9
11291 orf1ab G3676S
14408 orf1ab P314L
14676 orf1ab silent
15279 orf1ab silent
16176 orf1ab silent
18060 orf1ab silent
21764 S 21764_del6
21990 S 21990_del3
23063 S N501Y
23271 S A570D
23403 S D614G
23604 S P681H
23709 S T716I
24506 S S982A
24914 S D1118H
27972 orf8 Q27stop
28048 orf8 R52I
28095 orf8 K68stop
28111 orf8 Y73C
28144 orf8 S84L
28245 orf8 L118V
28280 N D3
28281 N D3
28282 N D3
28881 N R203K
28882 N R203K
28883 N G204R
28977 N S235F
48
50
50
50
50
50
50
50
50
50
50
91
50
50
50
50
50
50
50
50
50
50
50
50
50
50
50
59
67
70
77
93
88
88
89
48
47
47
52
VSP9977−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

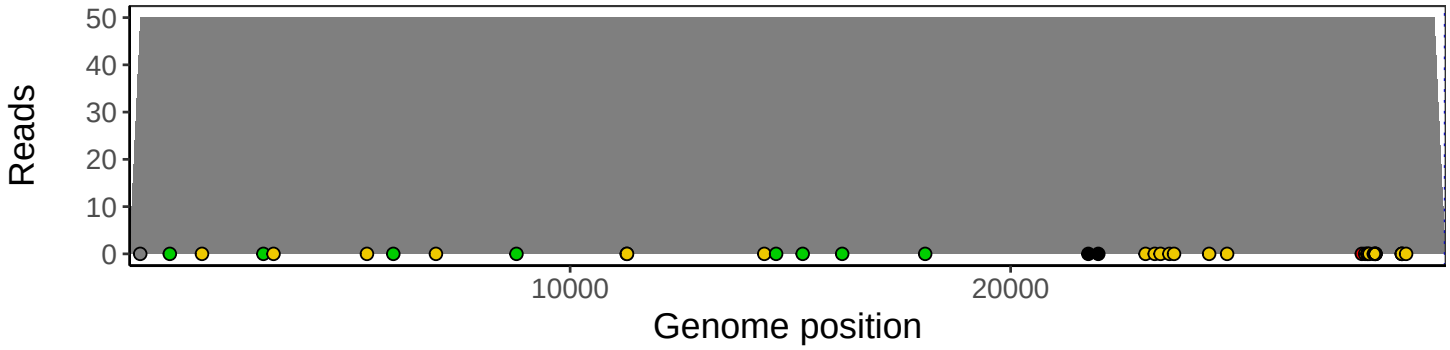
Analyses of individual experiments and composite results.

VSP9977-1 | 2020-12-11 | simulated | England_EPI_ISL_747519 | 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.