COVID-19 subject SouthAfrica_HLS-UCT-GS-1396-KRISP

2021-01-06

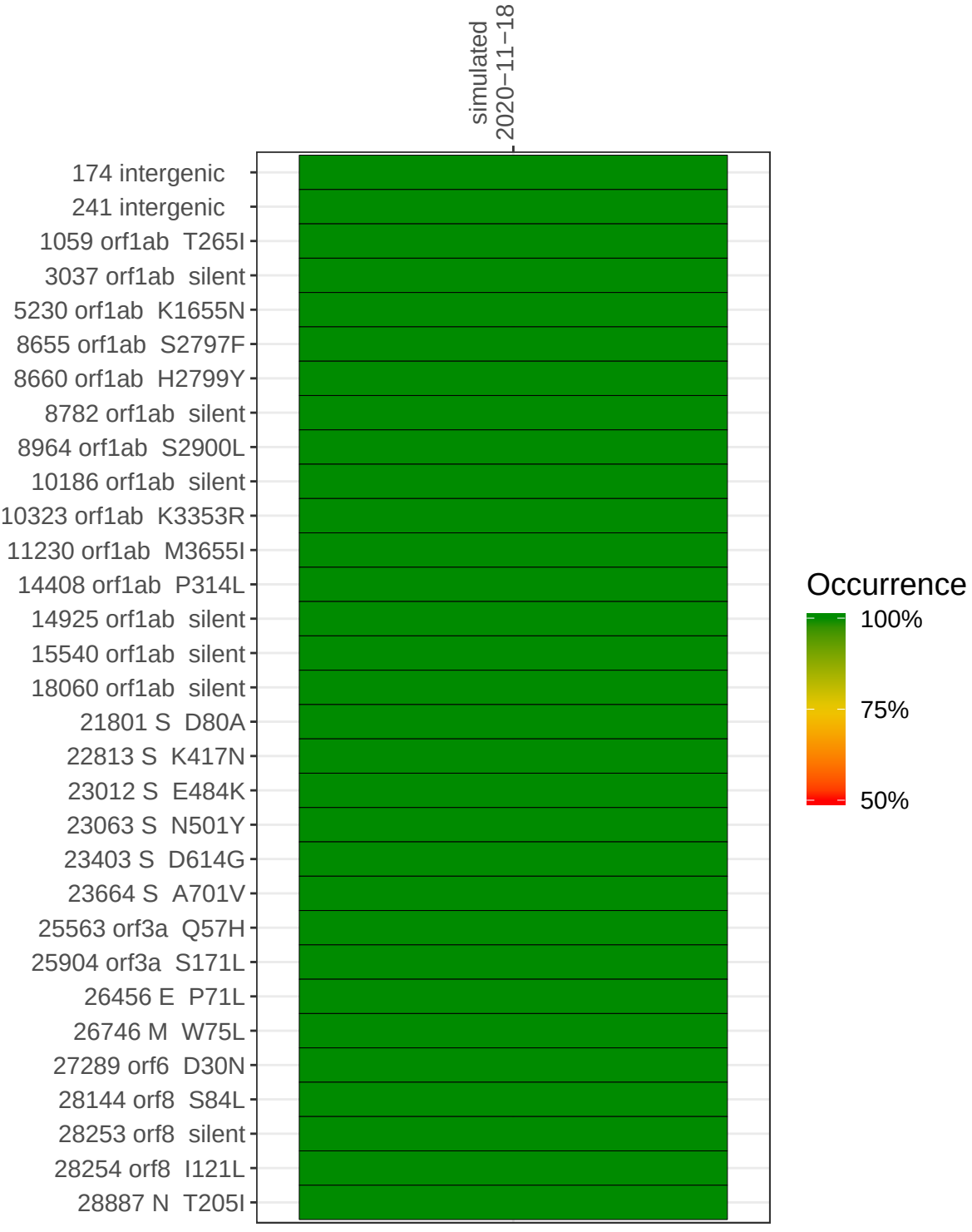
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. The code base for this analysis can be found (here).

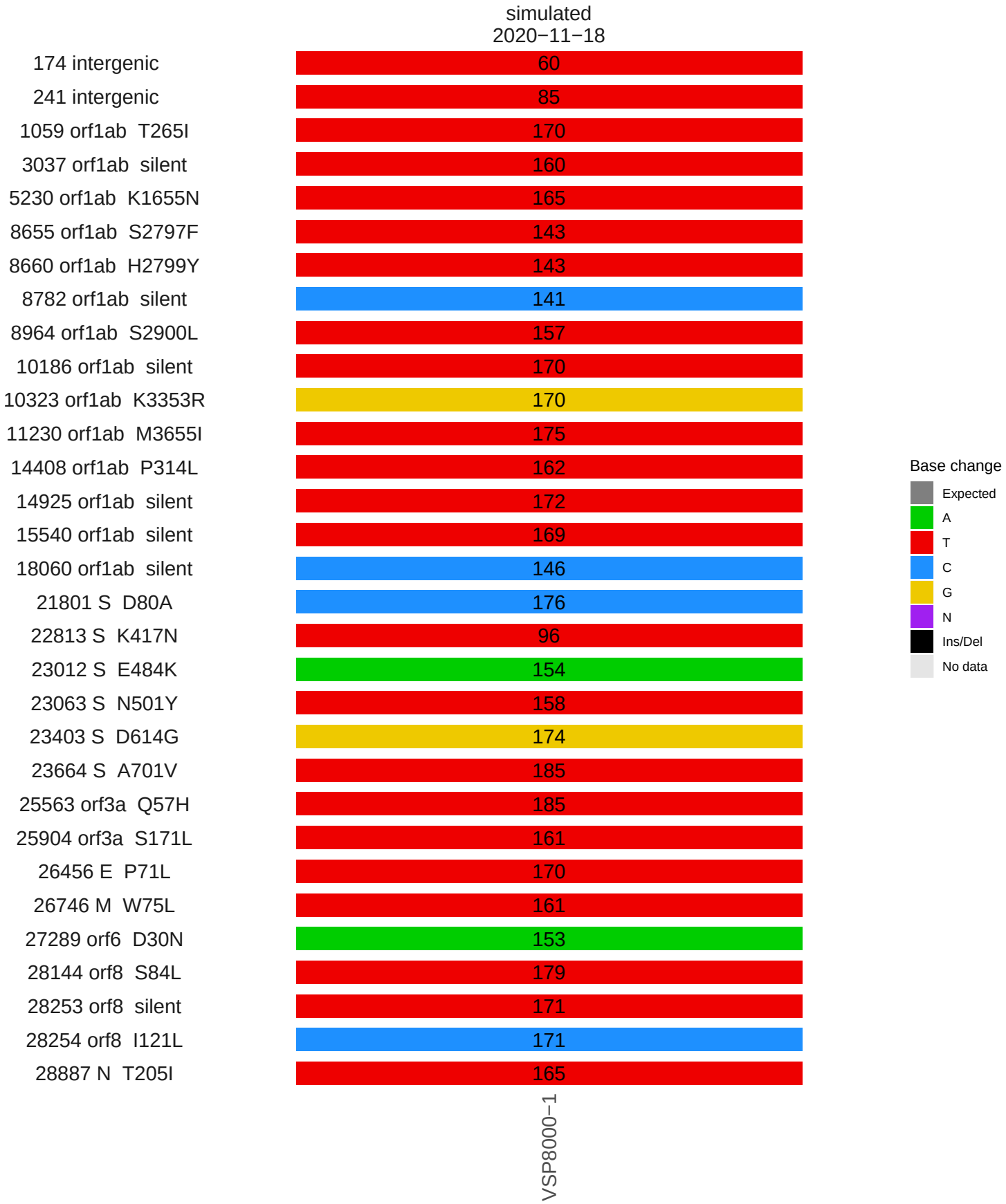
Table 1. Sample summary.

Experiment	Type	Input genomes	Sample type	Sample date	Largest contig (KD)	Reference read coverage	Reference read coverage (>= 5 reads)
VSP8000-1	single experiment	NA	simulated	2020-11-18	13.69	99.4%	99.0%

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 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.

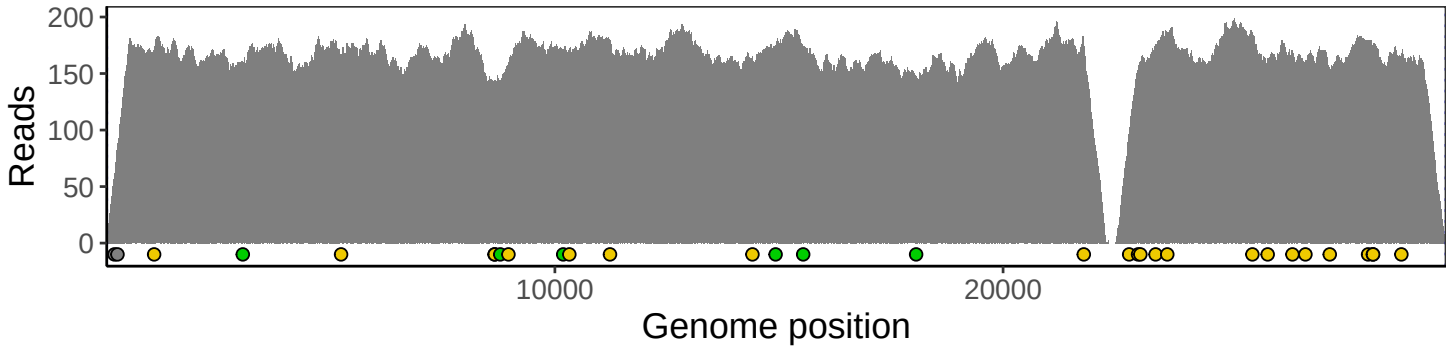



simulated
2020−11−18
174 intergenic
241 intergenic
1059 orf1ab T265I
3037 orf1ab silent
5230 orf1ab K1655N
8655 orf1ab S2797F
8660 orf1ab H2799Y
8782 orf1ab silent
8964 orf1ab S2900L
10186 orf1ab silent
10323 orf1ab K3353R
11230 orf1ab M3655I
14408 orf1ab P314L
14925 orf1ab silent
15540 orf1ab silent
18060 orf1ab silent
21801 S D80A
22813 S K417N
23012 S E484K
23063 S N501Y
23403 S D614G
23664 S A701V
25563 orf3a Q57H
25904 orf3a S171L
26456 E P71L
26746 M W75L
27289 orf6 D30N
28144 orf8 S84L
28253 orf8 silent
28254 orf8 I121L
28887 N T205I
60
85
170
160
165
143
143
141
157
170
170
175
162
172
169
146
176
96
154
158
174
185
185
161
170
161
153
179
171
171
165
VSP8000−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

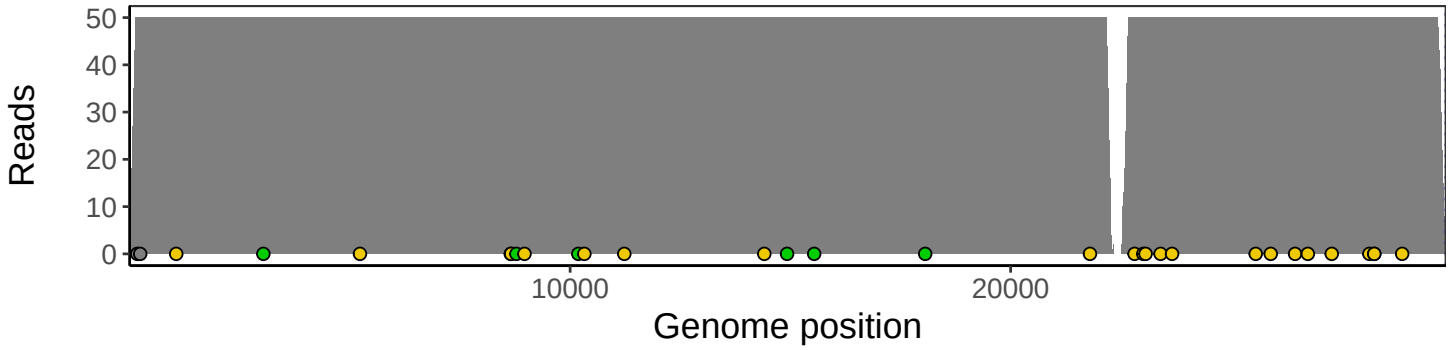
Analyses of individual experiments and composite results.

VSP8000-1 | 2020-11-18 | simulated | SouthAfrica_HLS-UCT-GS-1396-KRISP | 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.