COVID-19 subject CRD4J

2021-01-08

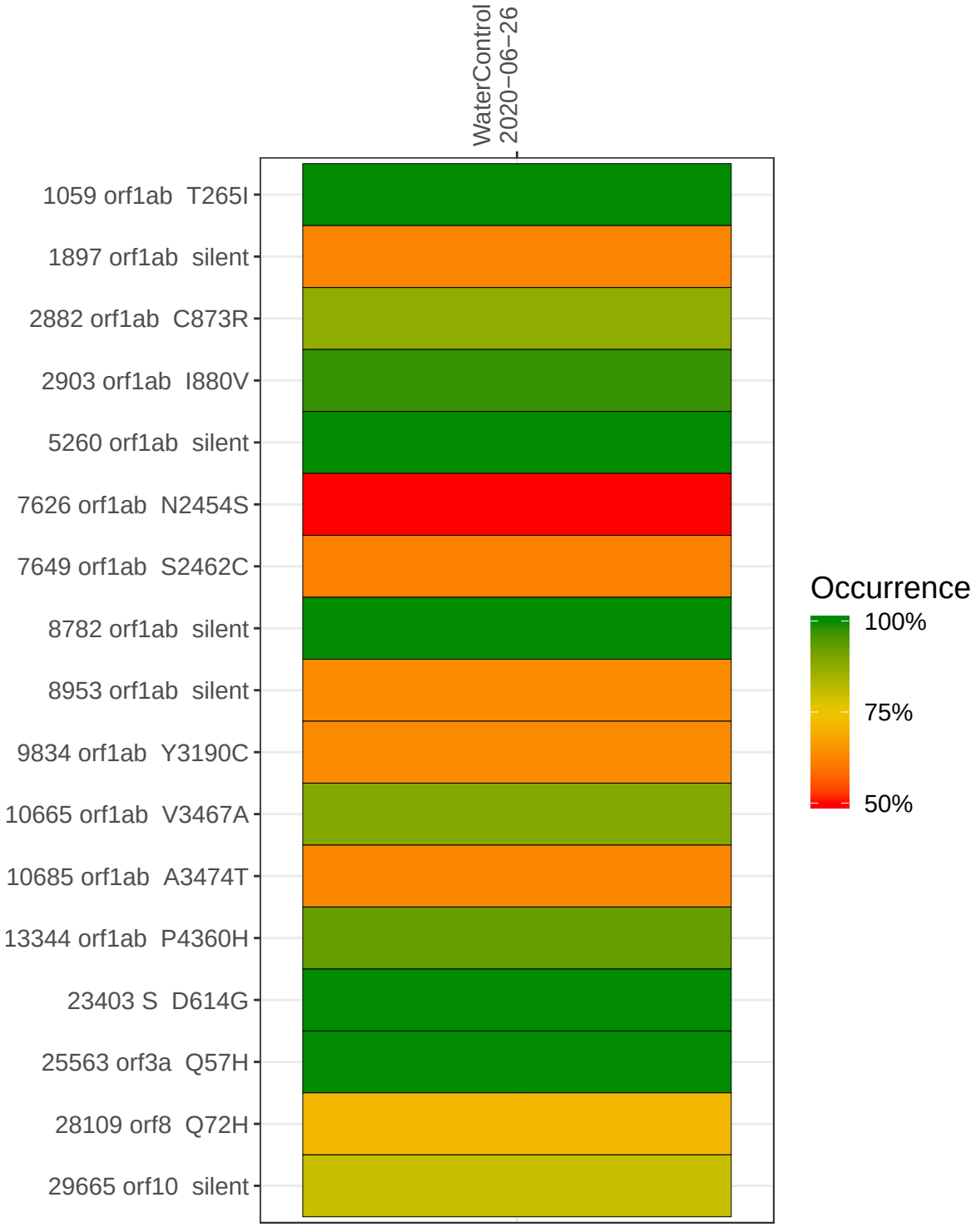
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. The code base for this analysis can be found (here).

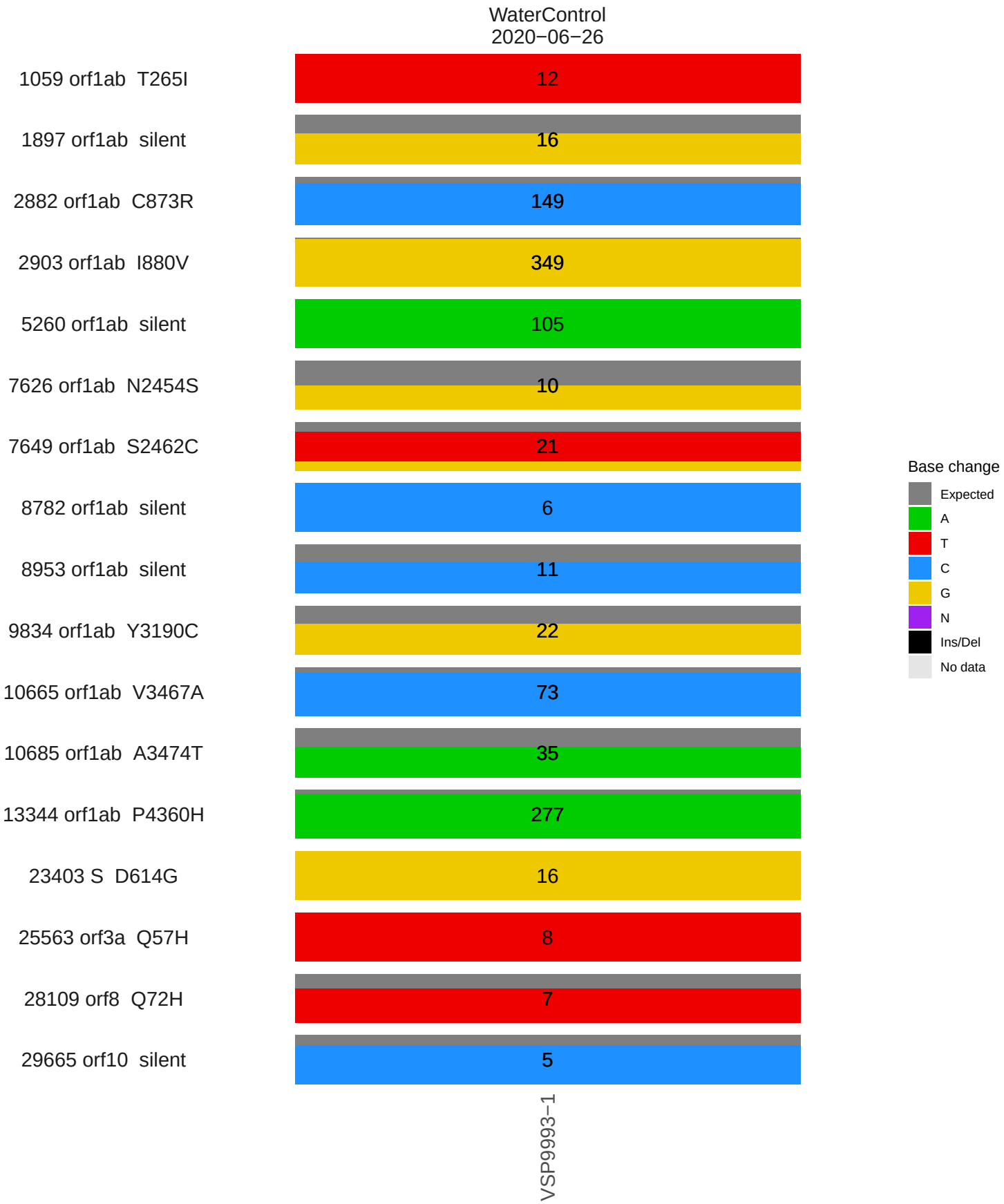
Table 1. Sample summary.

Experiment	Type	Input genomes	Sample type	Sample date	Largest contig (KD)	Reference read coverage	Reference read coverage (>= 5 reads)
VSP9993-1	single experiment	NA	WaterControl	2020-06-26	1.85	100.0%	59.4%

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.

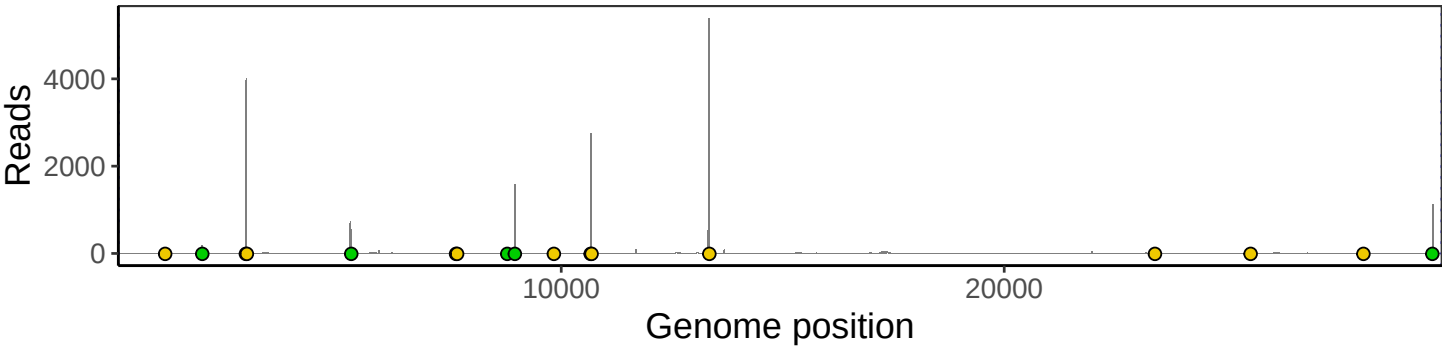



WaterControl
2020−06−26
1059 orf1ab T265I
12
1897 orf1ab silent
16
2882 orf1ab C873R
149
2903 orf1ab I880V
349
5260 orf1ab silent
105
7626 orf1ab N2454S
10
7649 orf1ab S2462C
21
8782 orf1ab silent
6
8953 orf1ab silent
11
9834 orf1ab Y3190C
22
10665 orf1ab V3467A
73
10685 orf1ab A3474T
35
13344 orf1ab P4360H
277
23403 S D614G
16
25563 orf3a Q57H
8
28109 orf8 Q72H
7
29665 orf10 silent
5
VSP9993−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

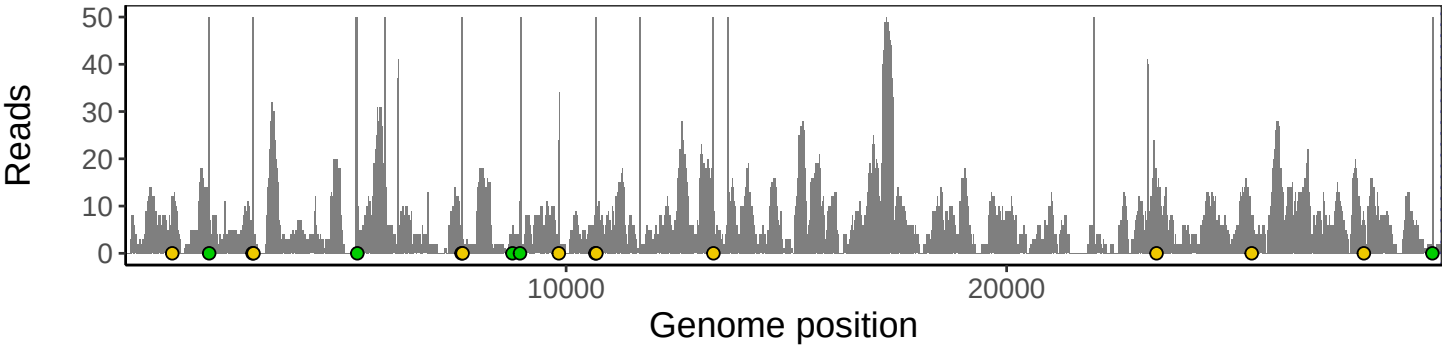
Analyses of individual experiments and composite results.

VSP9993-1 | 2020-06-26 | WaterControl | CRD4J | 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.