COVID-19 subject patient_C1

2021-04-17

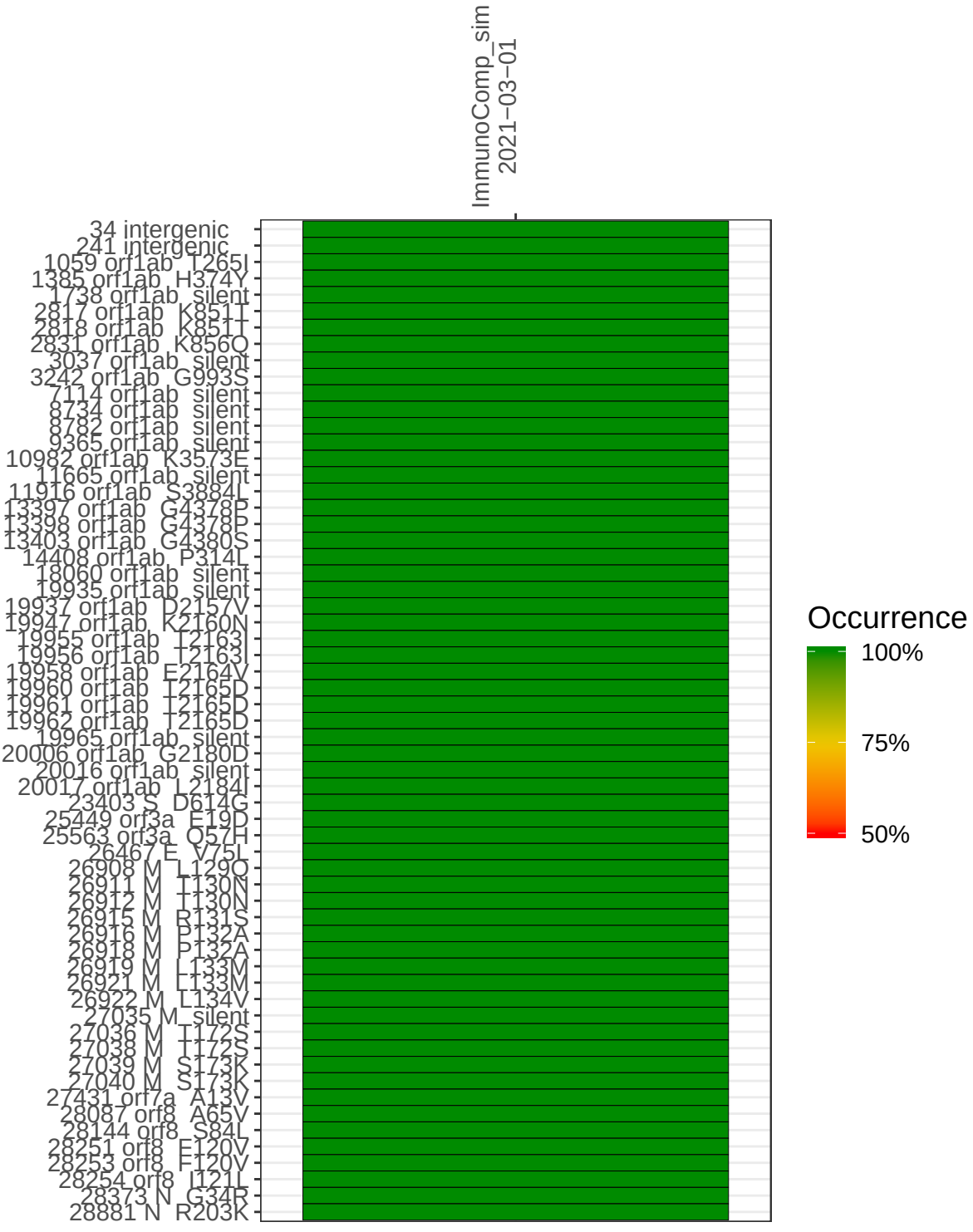
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

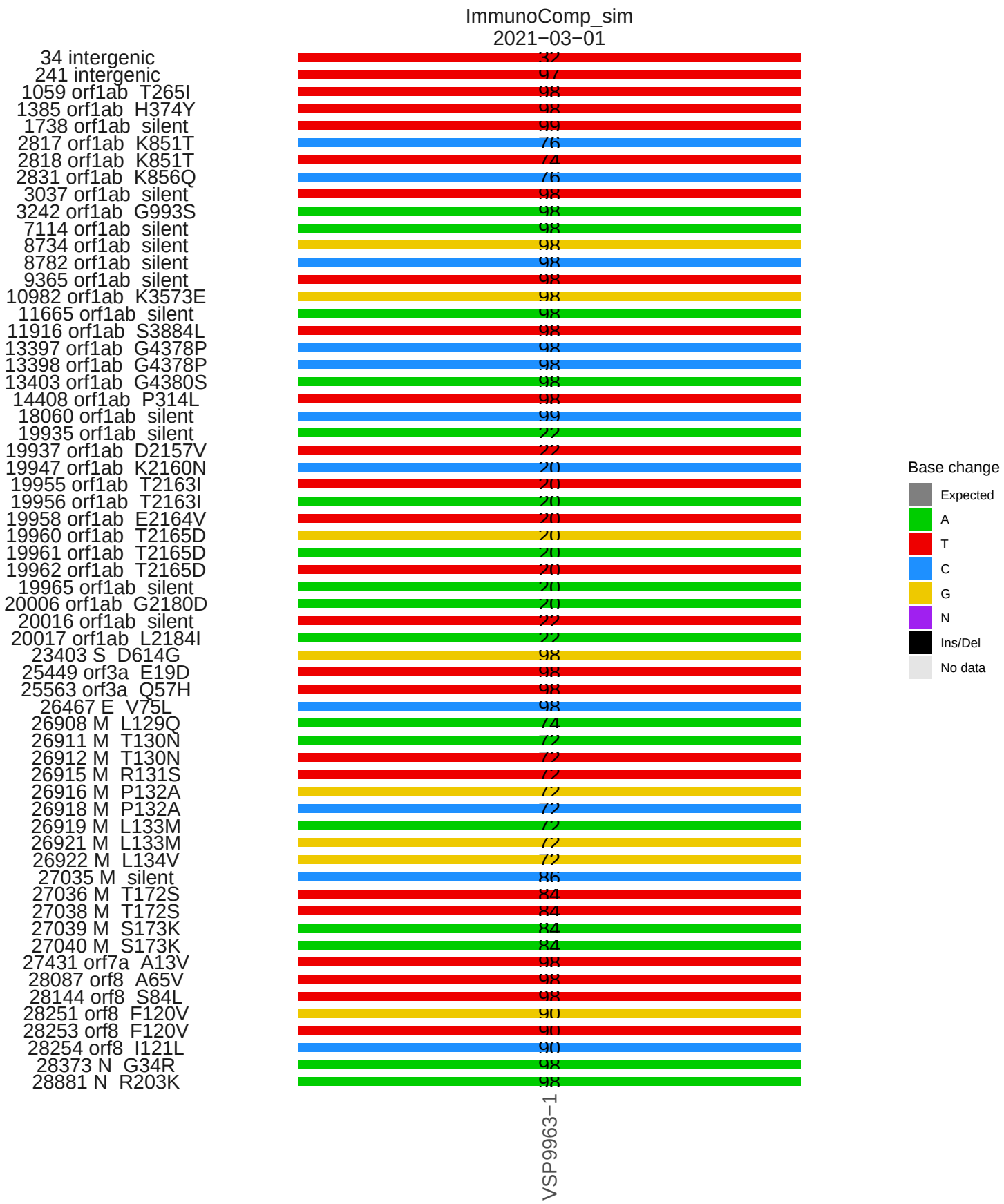
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP9963-1	single experiment	NA	ImmunoComp2021-03-01	2021-03-01	19.97	B.1.509	99.8%	99.8%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/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 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.

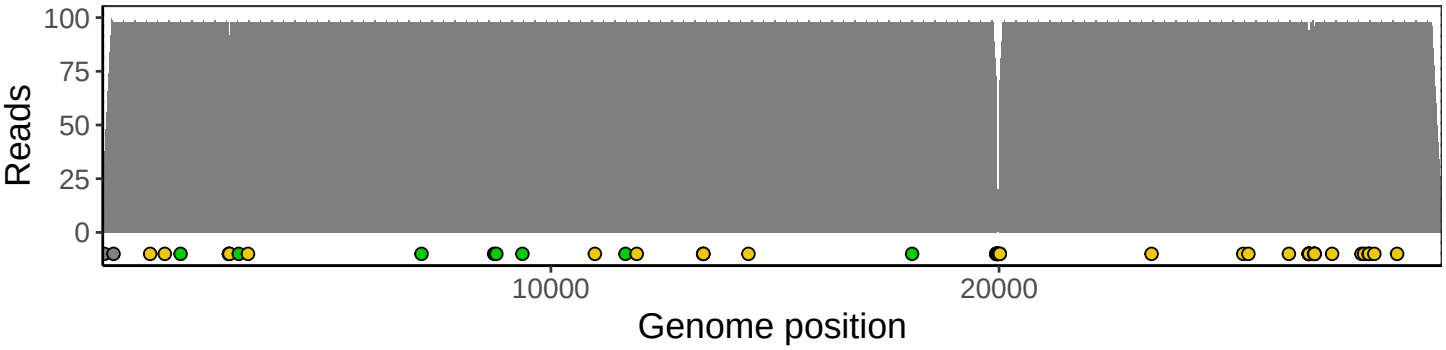



ImmunoComp_sim
2021−03−01
34 intergenic
241 intergenic
1059 orf1ab T265I
1385 orf1ab H374Y
1738 orf1ab silent
2817 orf1ab K851T
2818 orf1ab K851T
2831 orf1ab K856Q
3037 orf1ab silent
3242 orf1ab G993S
7114 orf1ab silent
8734 orf1ab silent
8782 orf1ab silent
9365 orf1ab silent
10982 orf1ab K3573E
11665 orf1ab silent
11916 orf1ab S3884L
13397 orf1ab G4378P
13398 orf1ab G4378P
13403 orf1ab G4380S
14408 orf1ab P314L
18060 orf1ab silent
19935 orf1ab silent
19937 orf1ab D2157V
19947 orf1ab K2160N
19955 orf1ab T2163I
19956 orf1ab T2163I
19958 orf1ab E2164V
19960 orf1ab T2165D
19961 orf1ab T2165D
19962 orf1ab T2165D
19965 orf1ab silent
20006 orf1ab G2180D
20016 orf1ab silent
20017 orf1ab L2184I
23403 S D614G
25449 orf3a E19D
25563 orf3a Q57H
26467 E V75L
26908 M L129Q
26911 M T130N
26912 M T130N
26915 M R131S
26916 M P132A
26918 M P132A
26919 M L133M
26921 M L133M
26922 M L134V
27035 M silent
27036 M T172S
27038 M T172S
27039 M S173K
27040 M S173K
27431 orf7a A13V
28087 orf8 A65V
28144 orf8 S84L
28251 orf8 F120V
28253 orf8 F120V
28254 orf8 I121L
28373 N G34R
28881 N R203K
VSP9963−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

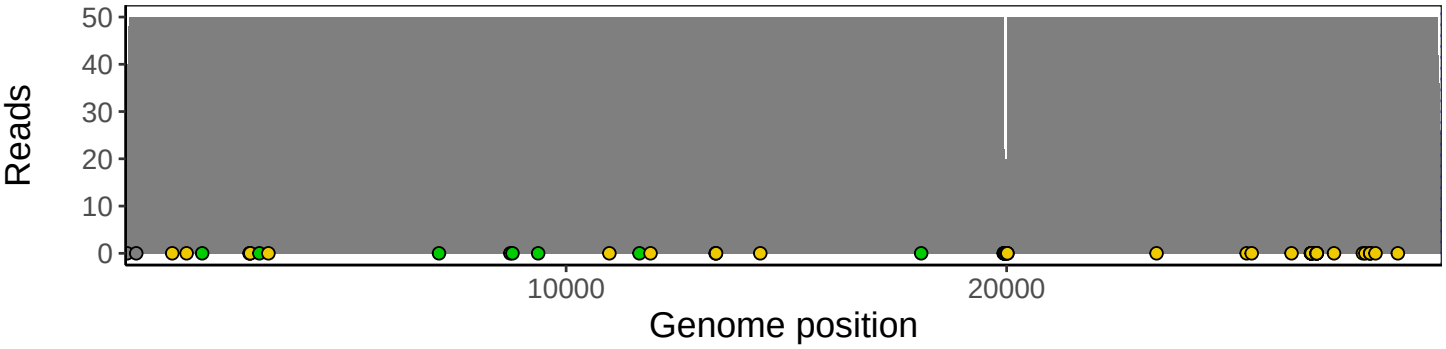
Analyses of individual experiments and composite results

VSP9963-1 | 2021-03-01 | ImmunoComp_sim | patient_C1 | genomes | single experiment

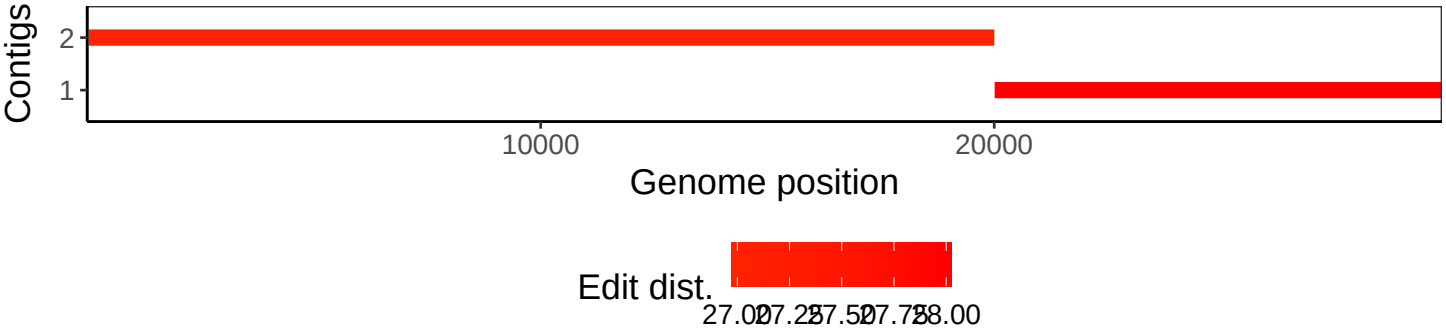
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	3.0.0
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