

COVID-19 subject patient__B1

2021-03-09

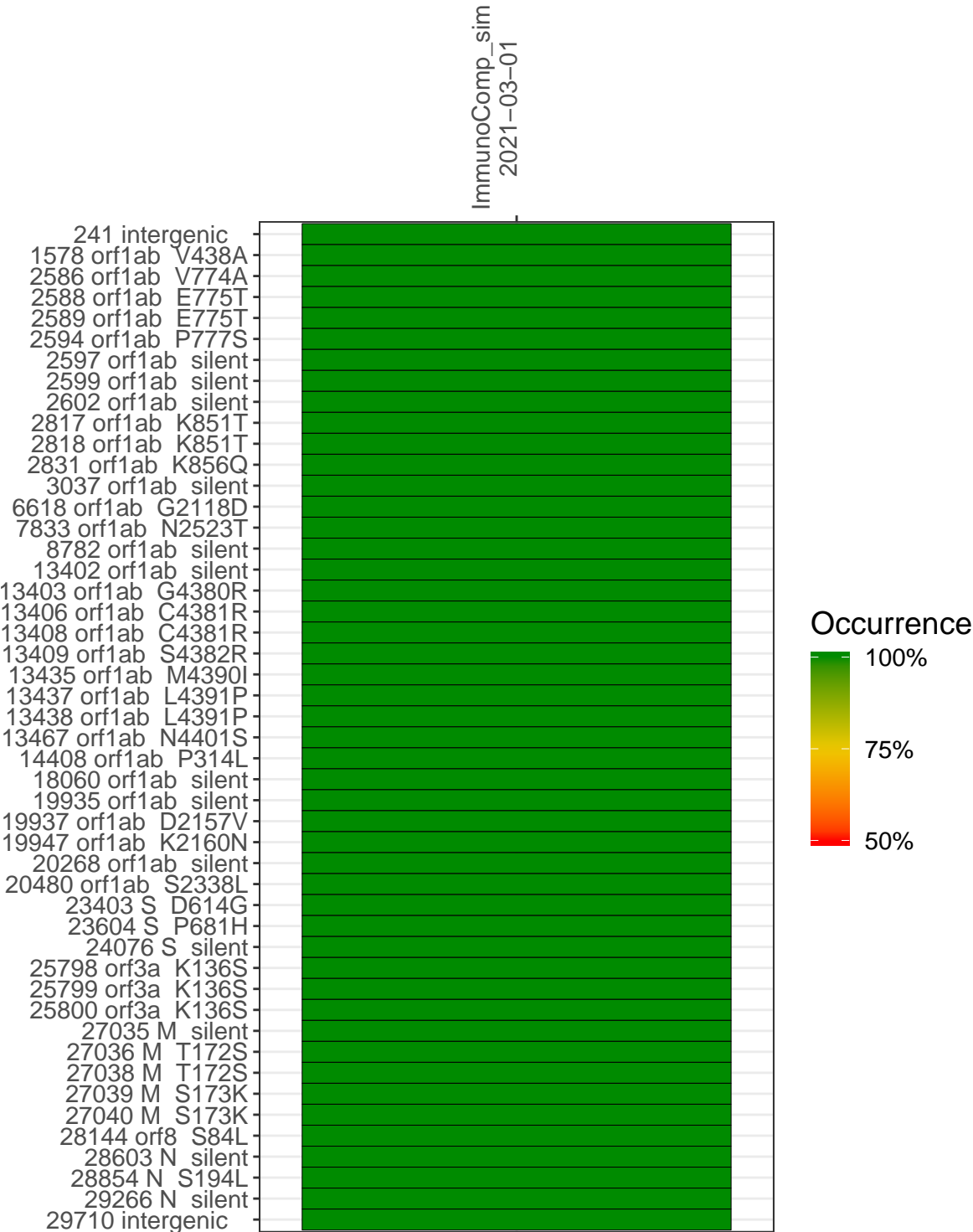
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP9964-1	single experiment	NA	ImmunoComp2021	2021-03-01	13.48	B.1.243	99.4%	99.4%

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 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

241 intergenic	97
1578 orf1ab V438A	98
2586 orf1ab V774A	82
2588 orf1ab E775T	80
2589 orf1ab E775T	80
2594 orf1ab P777S	80
2597 orf1ab silent	80
2599 orf1ab silent	80
2602 orf1ab silent	80
2817 orf1ab K851T	76
2818 orf1ab K851T	74
2831 orf1ab K856Q	76
3037 orf1ab silent	98
6618 orf1ab G2118D	98
7833 orf1ab N2523T	98
8782 orf1ab silent	98
13402 orf1ab silent	26
13403 orf1ab G4380R	26
13406 orf1ab C4381R	26
13408 orf1ab C4381R	26
13409 orf1ab S4382R	26
13435 orf1ab M4390I	26
13437 orf1ab L4391P	26
13438 orf1ab L4391P	26
13467 orf1ab N4401S	71
14408 orf1ab P314L	98
18060 orf1ab silent	98
19935 orf1ab silent	84
19937 orf1ab D2157V	84
19947 orf1ab K2160N	84
20268 orf1ab silent	98
20480 orf1ab S2338L	98
23403 S D614G	98
23604 S P681H	98
24076 S silent	98
25798 orf3a K136S	92
25799 orf3a K136S	92
25800 orf3a K136S	94
27035 M silent	79
27036 M T172S	79
27038 M T172S	79
27039 M S173K	79
27040 M S173K	79
28144 orf8 S84L	98
28603 N silent	98
28854 N S194L	98
29266 N silent	98
29710 intergenic	78

Base change

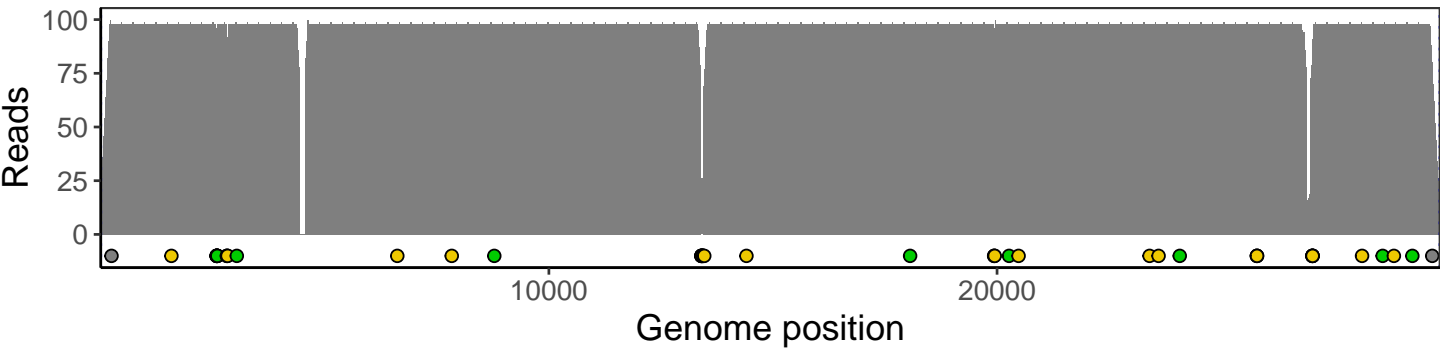
Expected
A
T
C
G
N
Ins/Del
No data

VSP9964-1

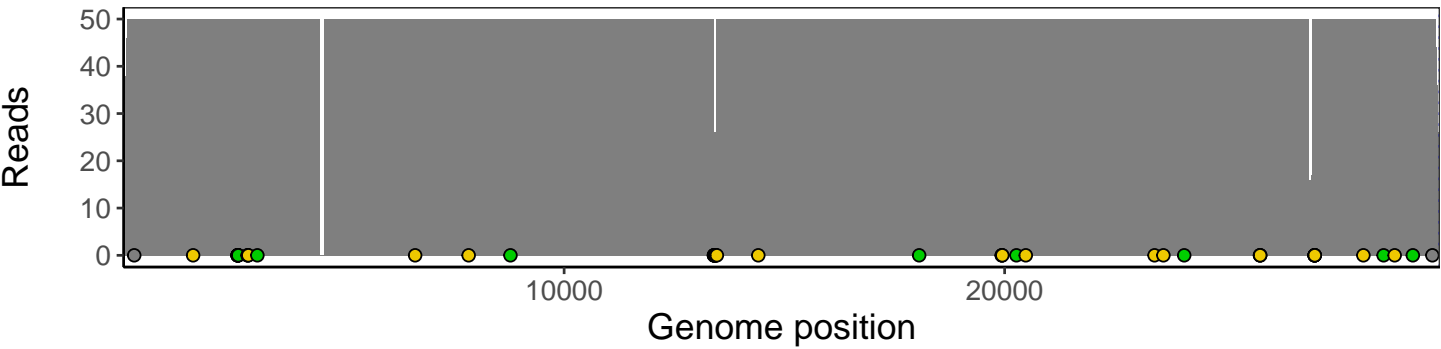
Analyses of individual experiments and composite results

VSP9964-1 | 2021-03-01 | ImmunoComp_sim | patient_B1 | genomes | single experiment

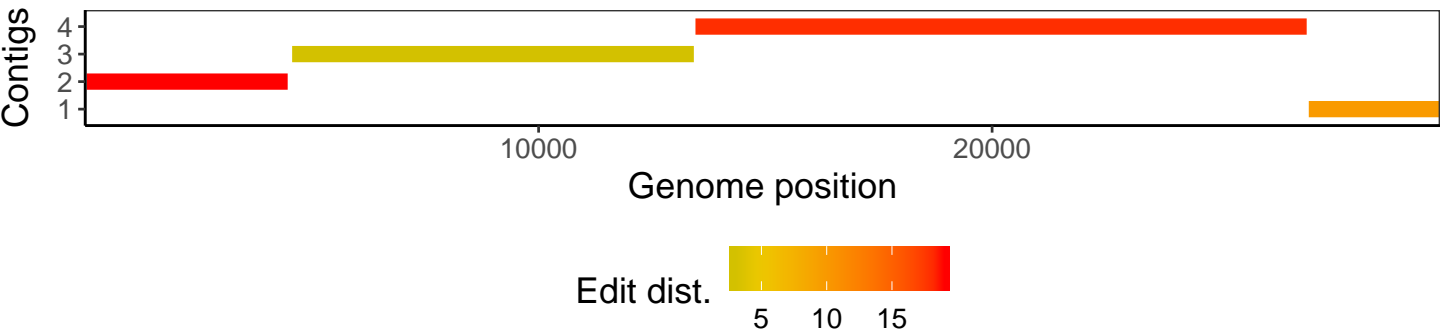
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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 htlib 1.10
bcftools	1.10.2-34-g1a12af0-dirty Using htlib 1.10.2-57-gf58a6f3
pangolin	2.3.3
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