

COVID-19 subject 263

2021-01-31

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.

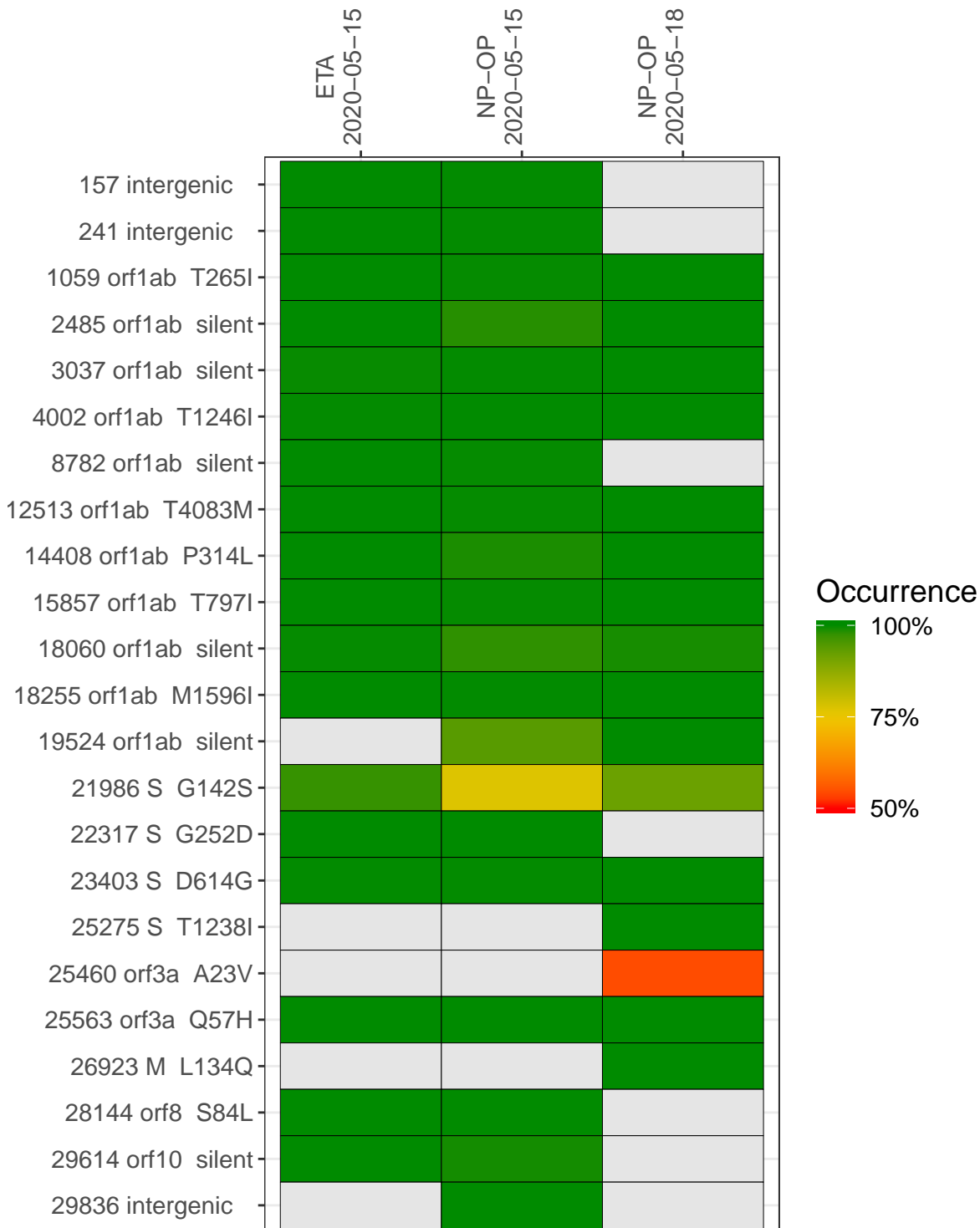
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with $> 90\%$ sequence coverage.

Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (≥ 5 reads)
VSP0142-1	single experiment	NA	ETA	2020-05-15	29.81	B.1.368	99.7%	99.7%
VSP0143-1	single experiment	172000	NP-OP	2020-05-15	29.87	B.1.368	99.8%	99.8%
VSP0148-1	single experiment	NA	NP-OP	2020-05-18	8.68	B.1.368	90.7%	90.2%

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.



	ETA 2020-05-15	NP-OP 2020-05-15	NP-OP 2020-05-18
157 intergenic	1318	6564	0
241 intergenic	1381	7266	0
1059 orf1ab T265I	775	5483	76
2485 orf1ab silent	731	5942	60
3037 orf1ab silent	1187	4548	158
4002 orf1ab T1246I	1475	7625	271
8782 orf1ab silent	933	5783	0
12513 orf1ab T4083M	1764	5937	395
14408 orf1ab P314L	2115	8068	469
15857 orf1ab T797I	2110	9521	506
18060 orf1ab silent	828	5667	153
18255 orf1ab M1596I	1177	5321	191
19524 orf1ab silent	263	4217	24
21986 S G142S	343	3051	74
22317 S G252D	30	941	12
23403 S D614G	3646	12215	523
25275 S T1238I	586	5440	53
25460 orf3a A23V	1418	8973	200
25563 orf3a Q57H	1288	7794	156
26923 M L134Q	1805	9752	105
28144 orf8 S84L	1519	11171	45
29614 orf10 silent	256	2618	0
29836 intergenic	1	308	0
	VSP0142-1	VSP0143-1	VSP0148-1

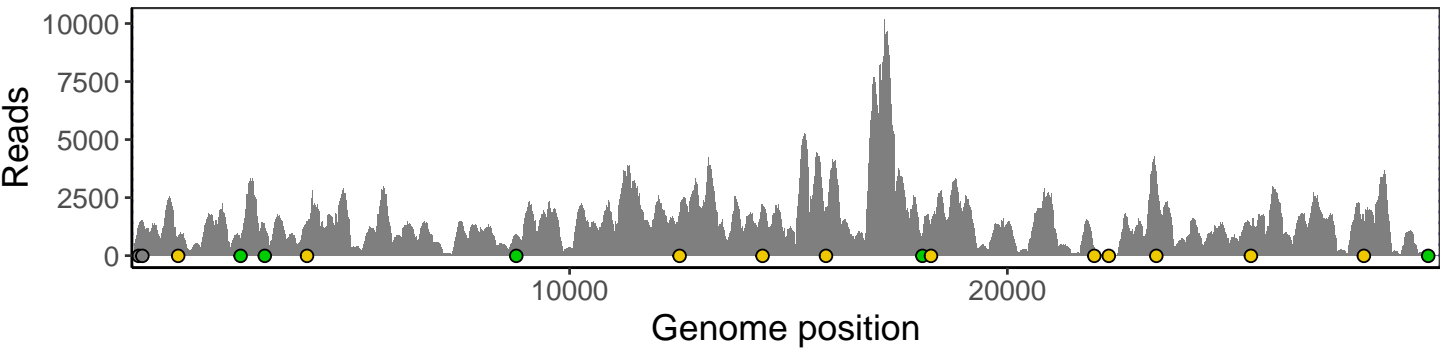
Base change

- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

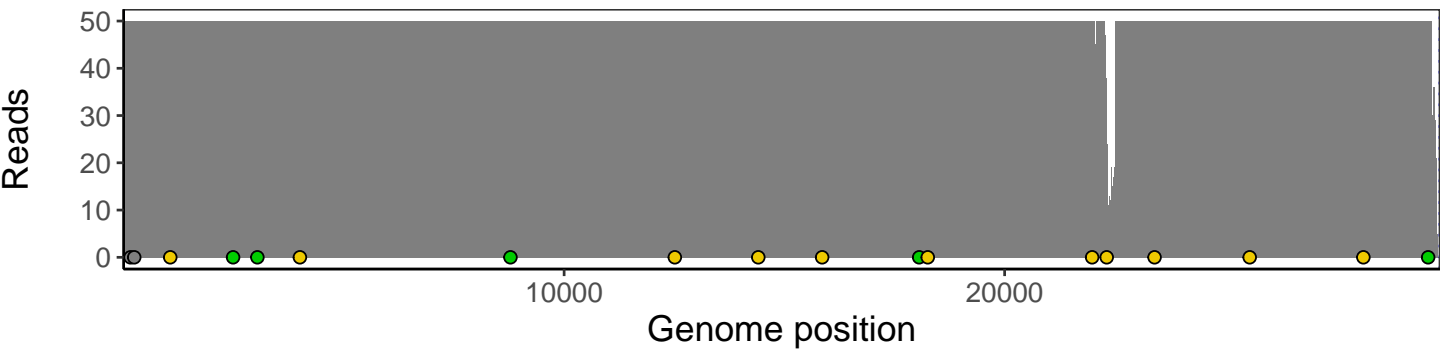
Analyses of individual experiments and composite results

VSP0142-1 | 2020-05-15 | ETA | 263e-q1 | genomes | single experiment

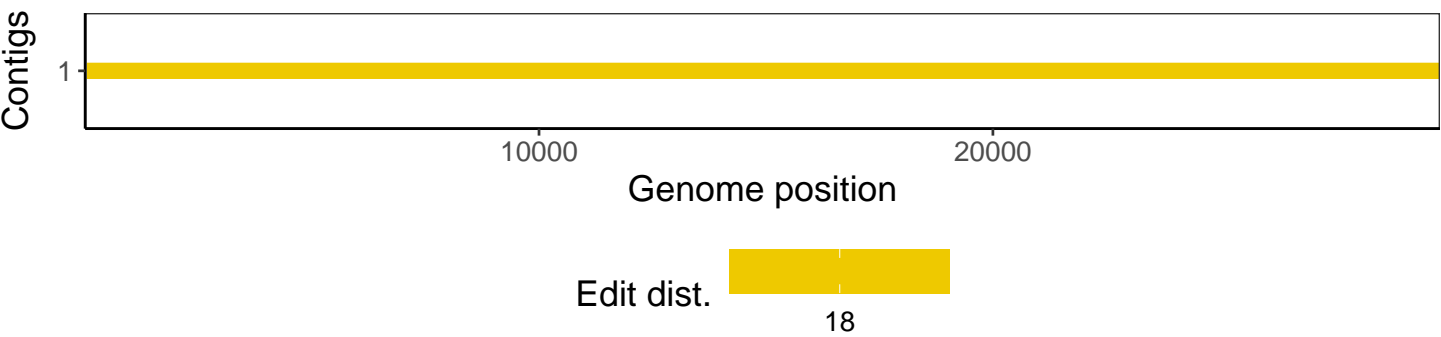
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



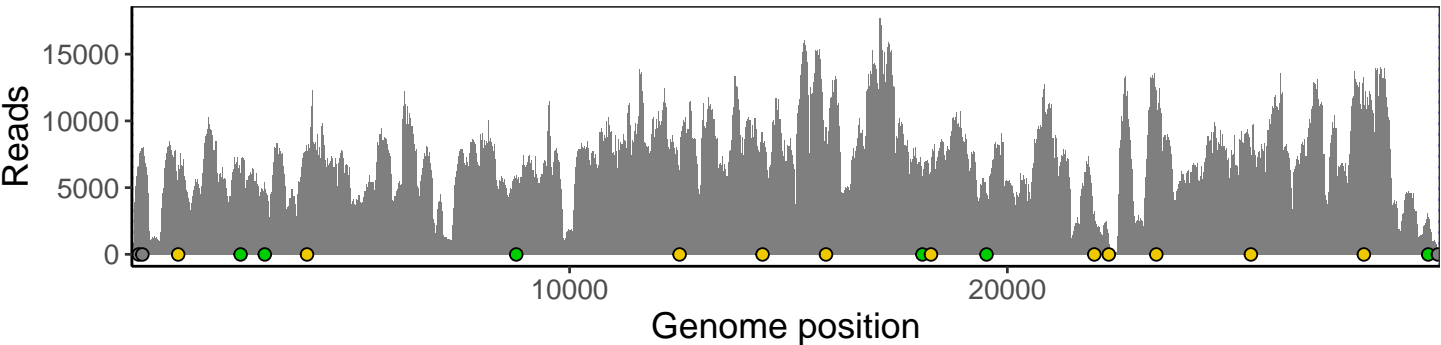
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



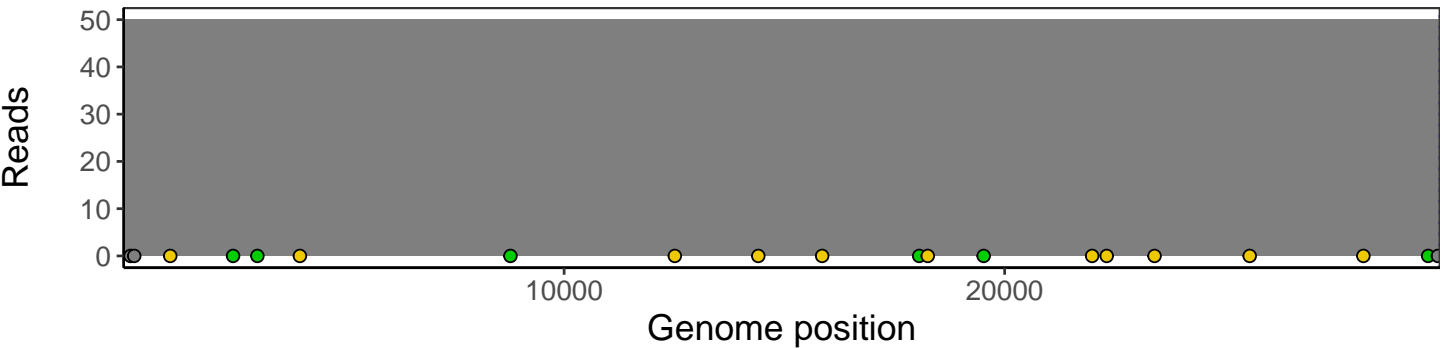
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



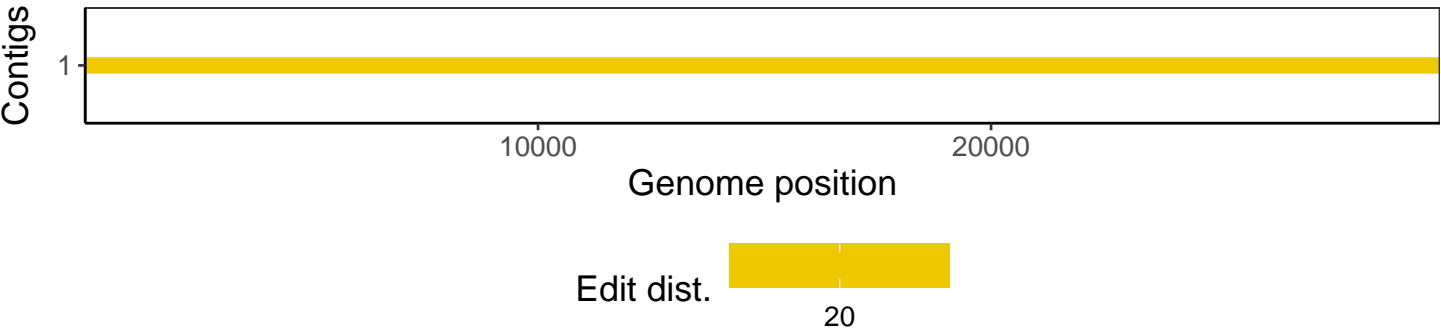
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



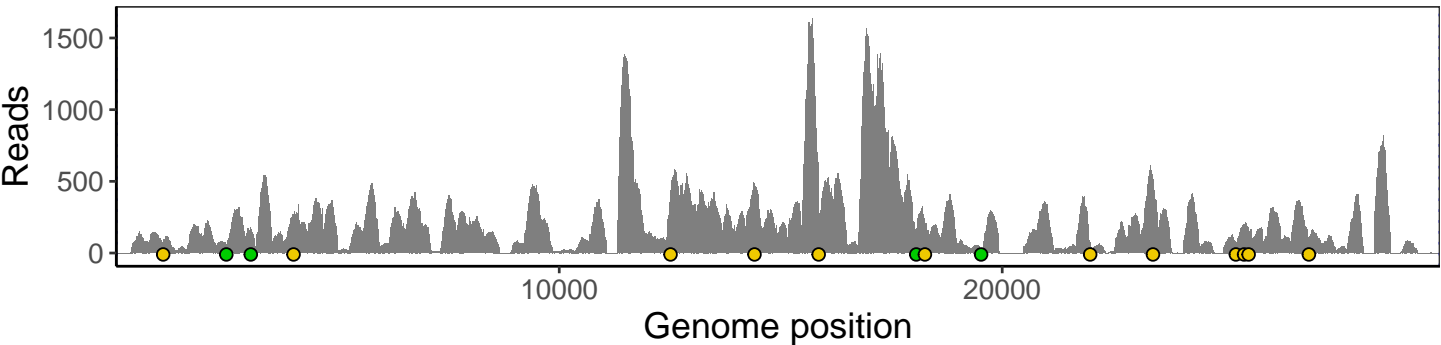
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.



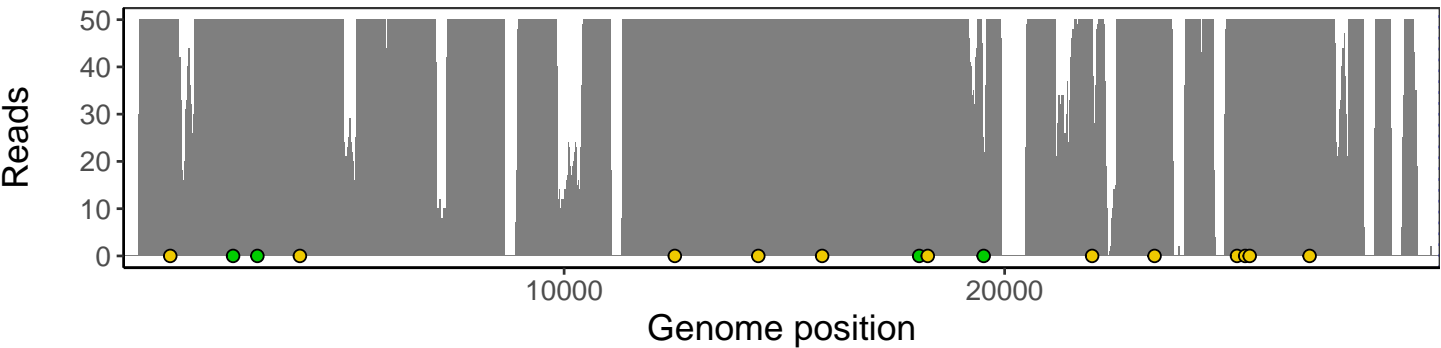
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.



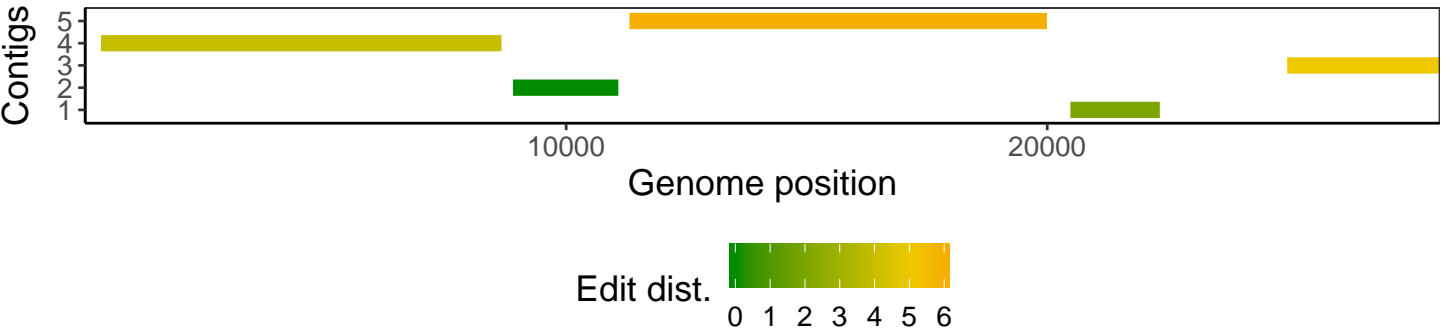
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	2.1.7
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