COVID-19 subject 424

2021-05-05

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)
VSP0517	composite	NA	NP-OP	2020-12-09	29.94	B.1.243	99.9%	99.8%
VSP0517-1	single experiment	NA	NP-OP	2020-12-09	29.85	B.1.243	99.9%	99.8%
VSP0517-2	single experiment	NA	NP-OP	2020-12-09	19.35	B.1.243	99.7%	99.7%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/Wuhan-Hu-1) 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.

NP−OP
2020−12−09
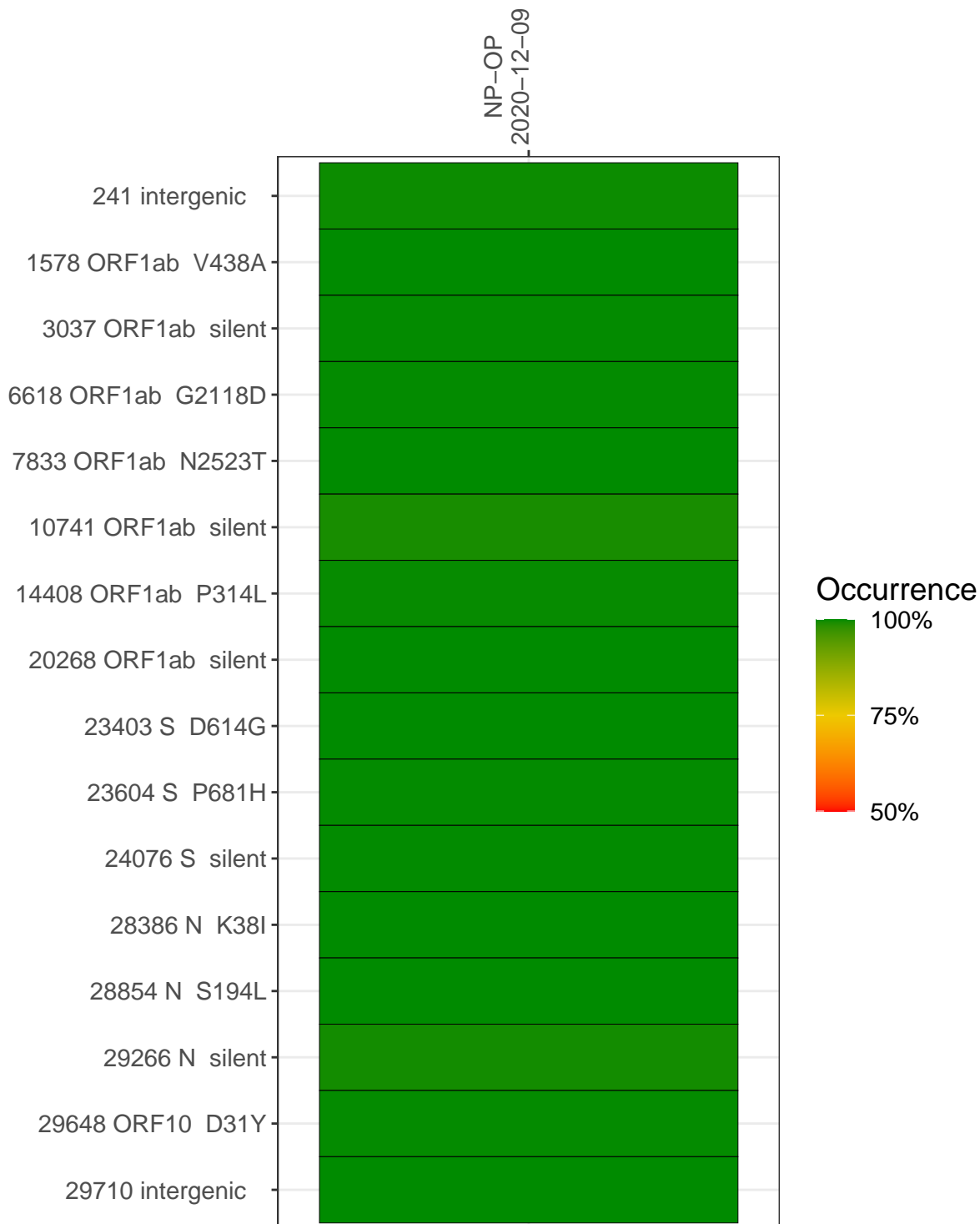
Mutation	VSP0517−1	VSP0517−2
241 intergenic	5498	2271
1578 ORF1ab V438A	1029	418
3037 ORF1ab silent	3531	1361
6618 ORF1ab G2118D	6225	2586
7833 ORF1ab N2523T	6909	2482
10741 ORF1ab silent	3027	1013
14408 ORF1ab P314L	7472	2653
20268 ORF1ab silent	1103	351
23403 S D614G	12841	4735
23604 S P681H	9285	3461
24076 S silent	1754	589
28386 N K38I	10836	3808
28854 N S194L	1612	592
29266 N silent	6189	2098
29648 ORF10 D31Y	2891	1056
29710 intergenic	725	306

Base change
Expected
A
T
C
G
N
Ins/Del
No data
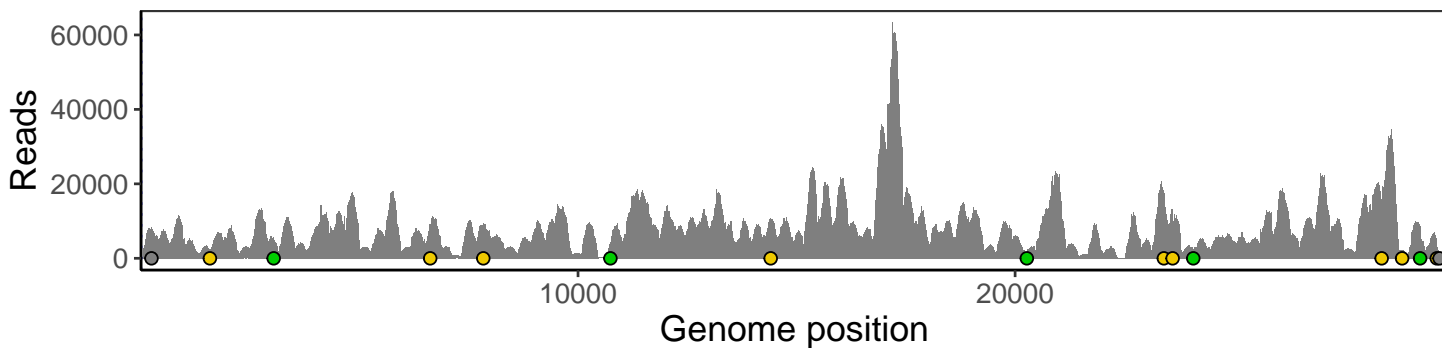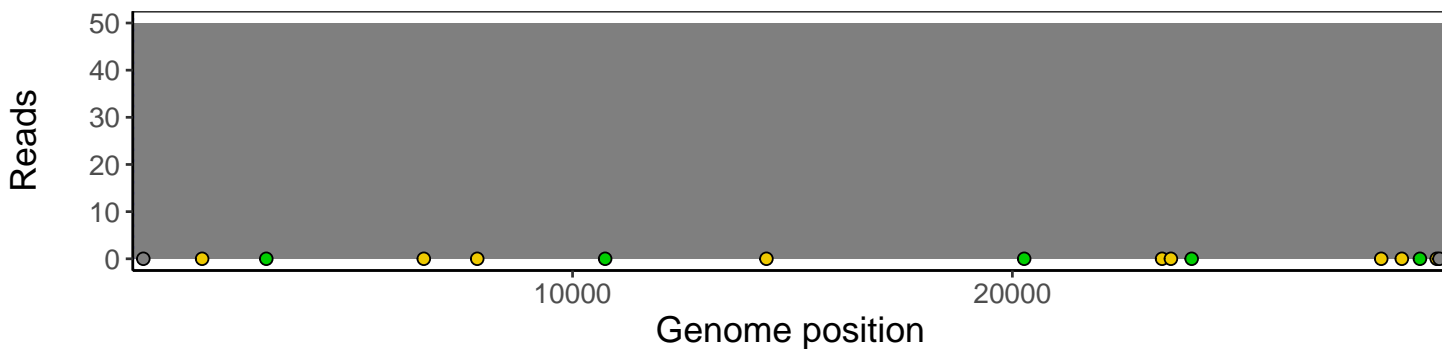
3

Analyses of individual experiments and composite results

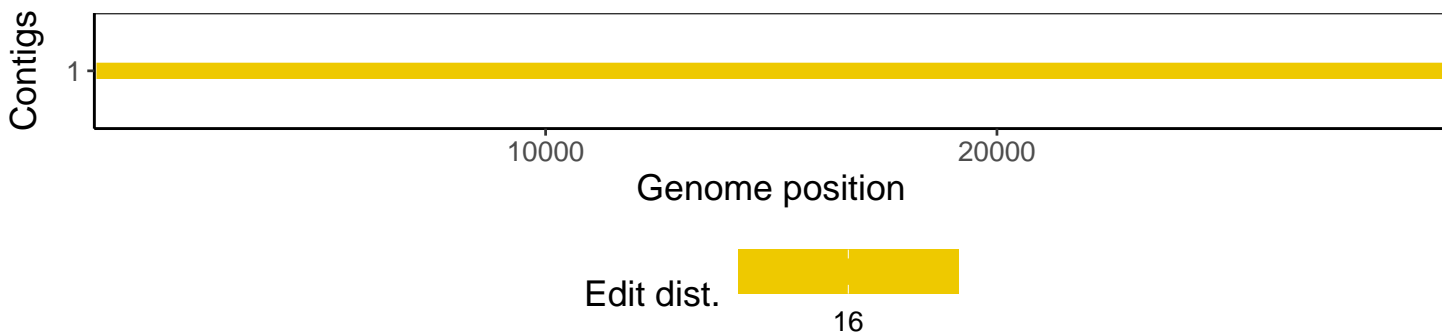## VSP0517 | 2020-12-09 | NP-OP | 424no | composite result

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.
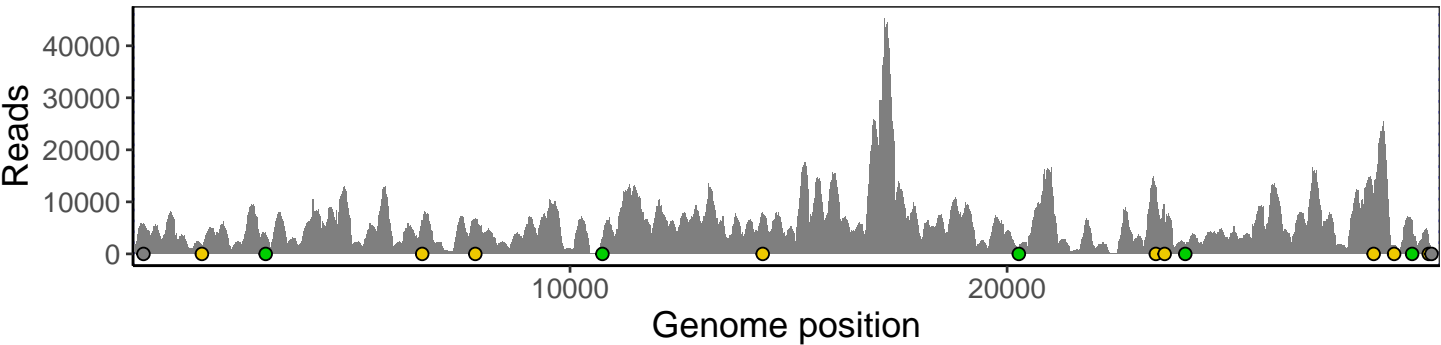

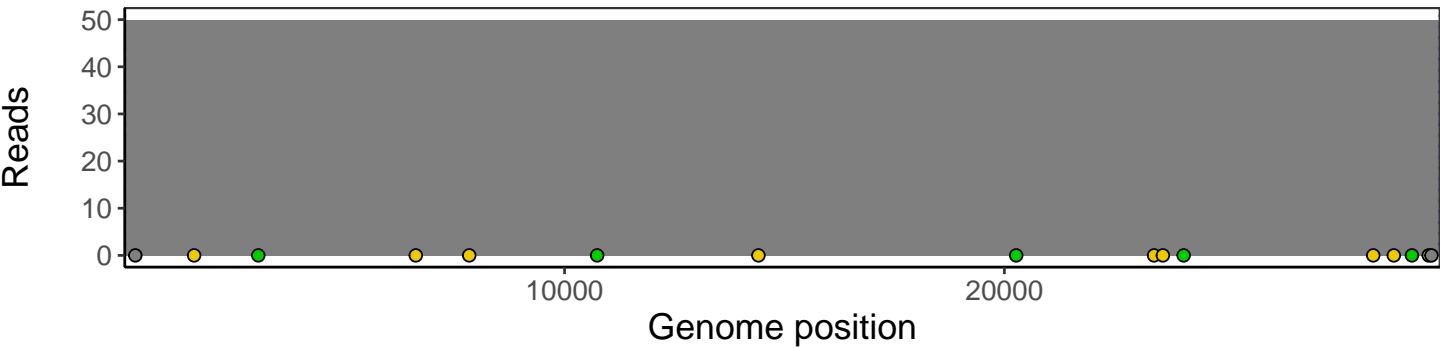Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

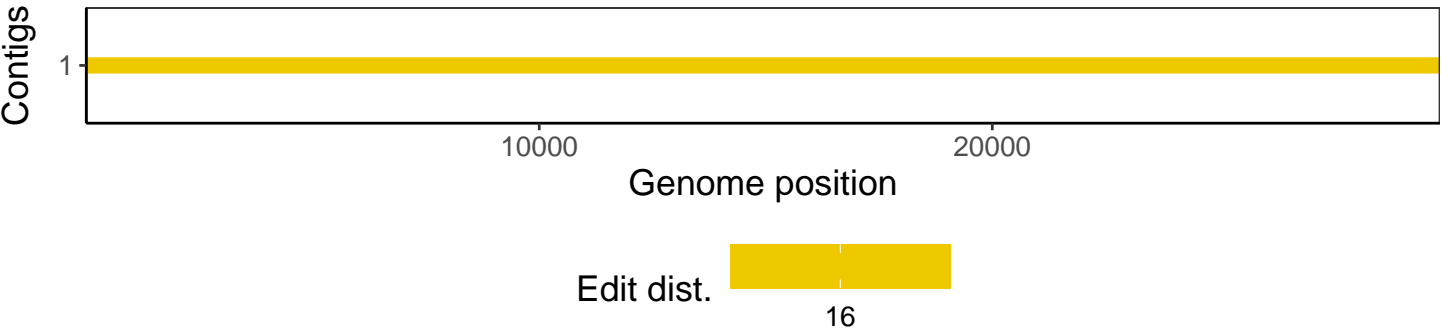**VSP0517-1 | 2020-12-09 | NP-OP | 424no | genomes | single experiment**

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.
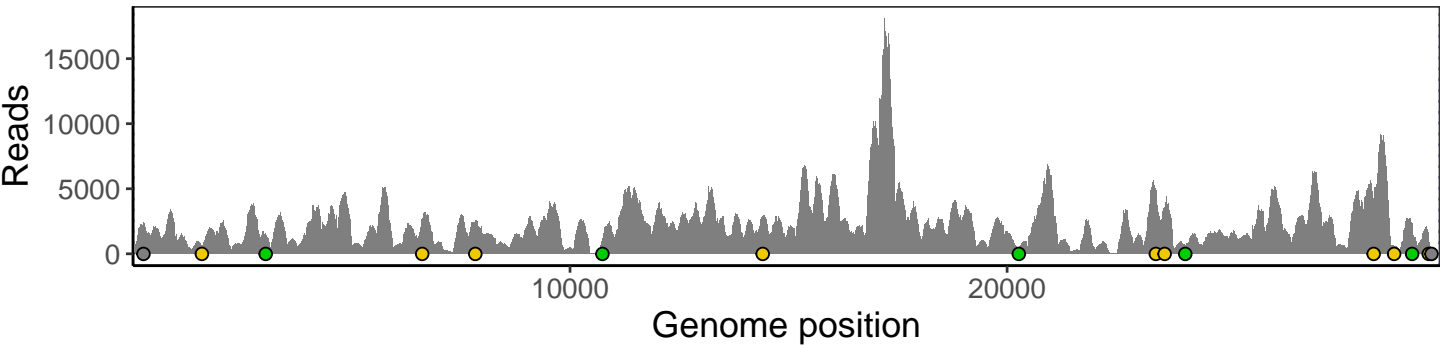

Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

VSP0517-2 | 2020-12-09 | NP-OP | 424no | 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.

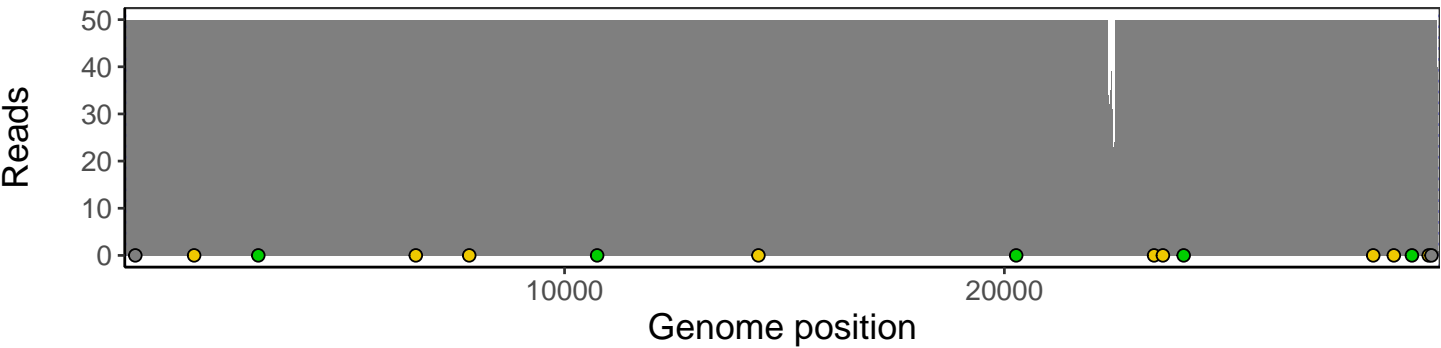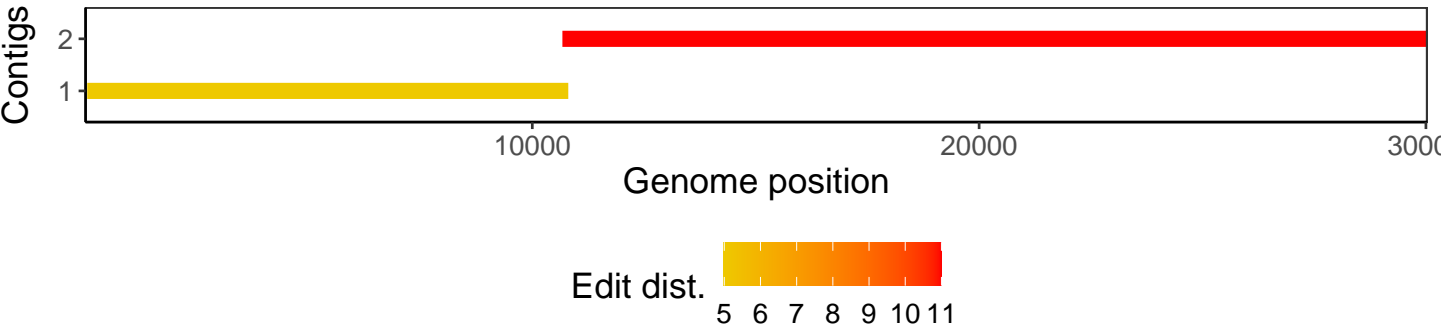Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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