COVID-19 subject 211-TCE

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)
VSP0184-1m	single experiment	NA	NP	2020-04-08	29.86	B.1	99.9%	99.9%
VSP0185-1m	single experiment	NA	OP	2020-04-13	29.88	B.1	99.9%	99.9%
VSP0186-1m	single experiment	NA	ETA	2020-04-13	29.96	B.1	99.9%	99.9%

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.
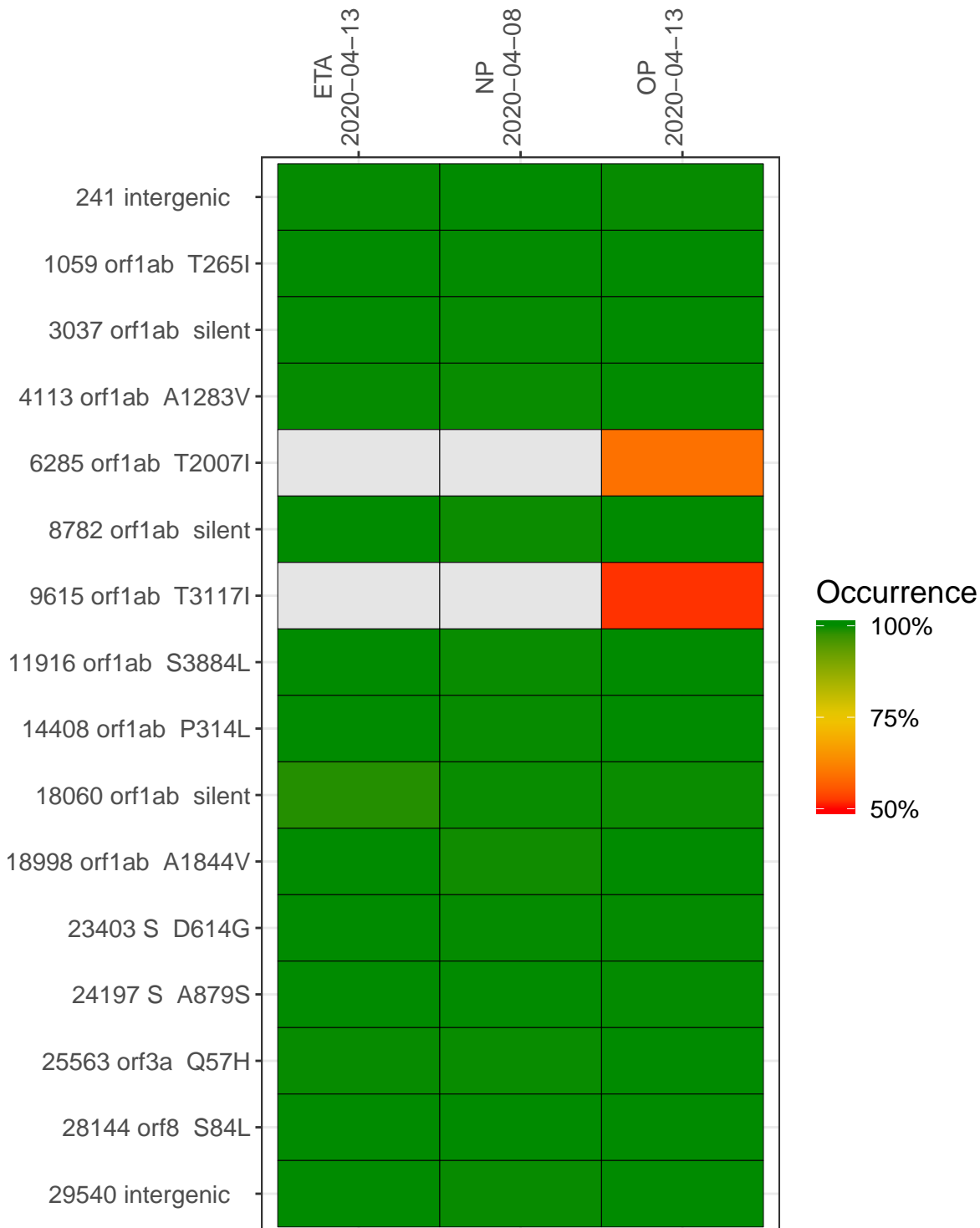
	ETA 2020−04−13	NP 2020−04−08	OP 2020−04−13
241 intergenic	1938	2184	2035
1059 orf1ab T265I	1314	4984	5385
3037 orf1ab silent	1199	2086	3406
4113 orf1ab A1283V	2514	3703	3527
6285 orf1ab T2007I	1869	3554	263
8782 orf1ab silent	886	1487	242
9615 orf1ab T3117I	667	123	1329
11916 orf1ab S3884L	1165	980	2590
14408 orf1ab P314L	1062	1416	3291
18060 orf1ab silent	1146	1082	3291
18998 orf1ab A1844V	1341	2846	217
23403 S D614G	4146	3933	3124
24197 S A879S	1106	4543	4675
25563 orf3a Q57H	1329	4824	5037
28144 orf8 S84L	1120	2515	791
29540 intergenic	760	1223	889
	VSP0186−1m	VSP0184−1m	VSP0185−1m

Base change
- Expected (grey)
- A (green)
- T (red)
- C (blue)
- G (yellow)
- N (purple)
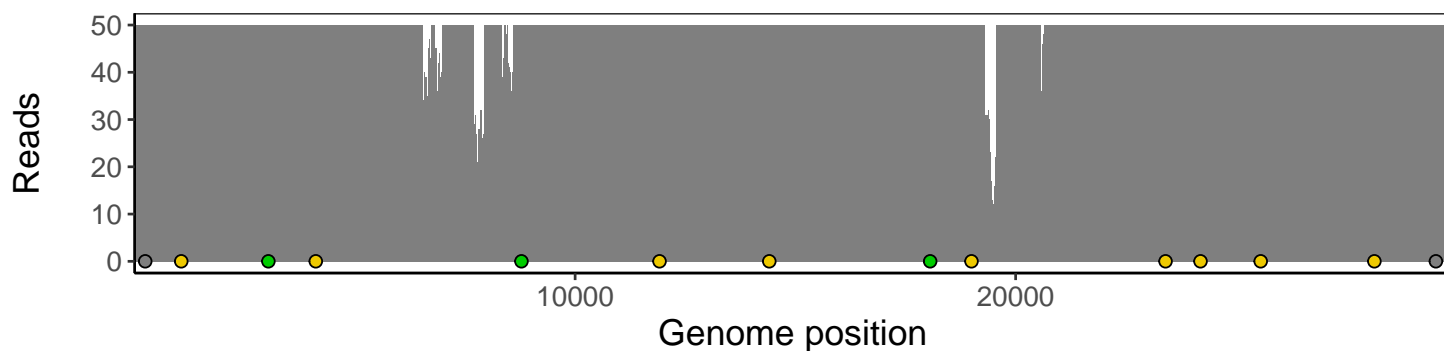- Ins/Del (black)
- No data (light grey)

3

Analyses of individual experiments and composite results

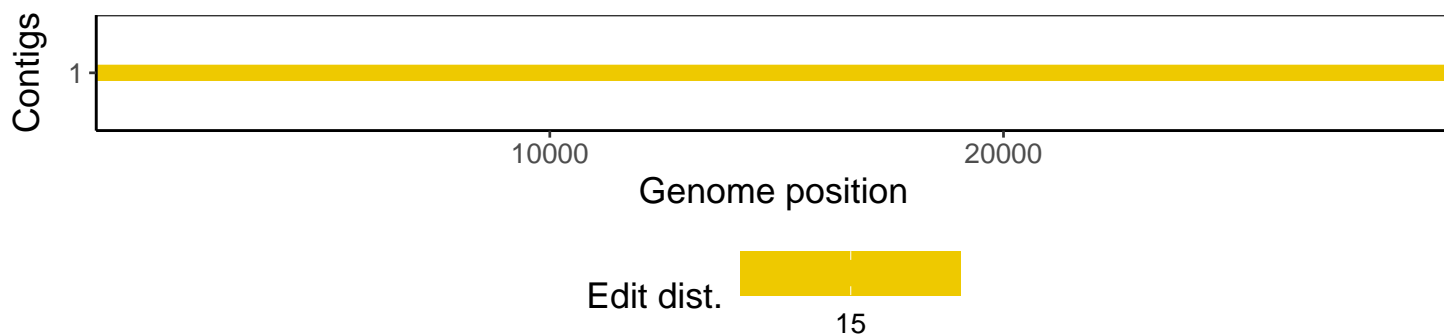## VSP0184-1m | 2020-04-08 | NP | 5 | genomes | single experiment

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.
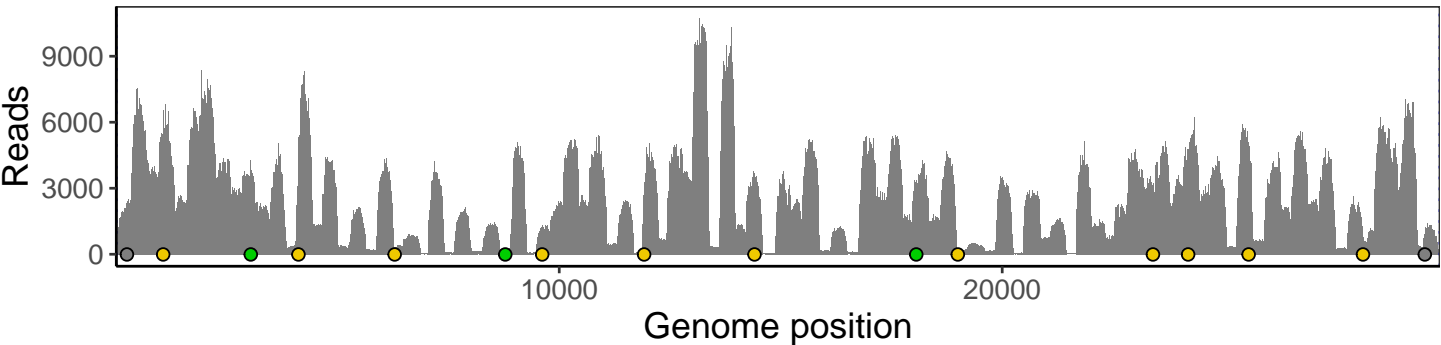

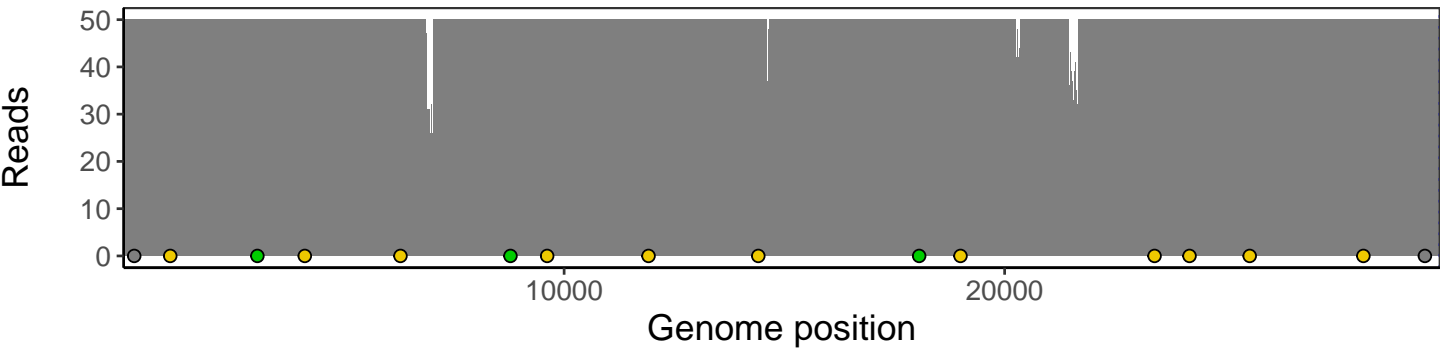Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

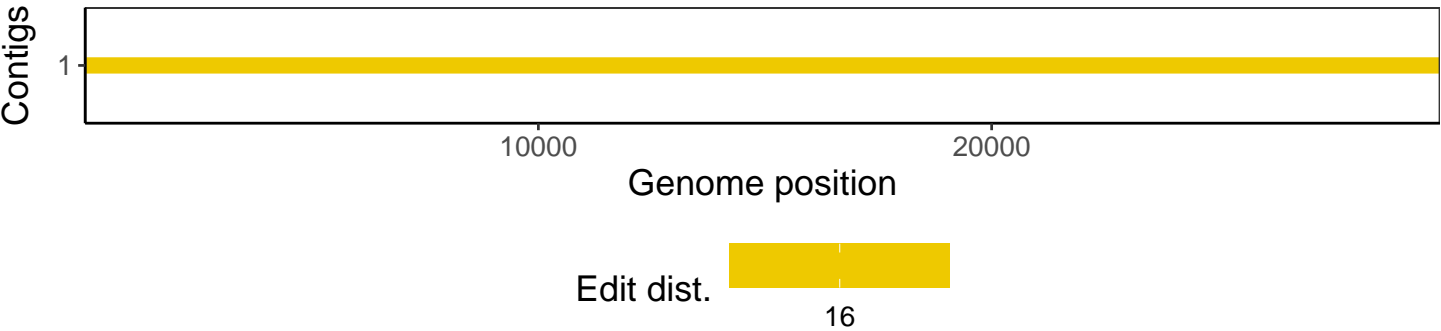**VSP0185-1m | 2020-04-13 | OP | 6 | genomes | single experiment**

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.
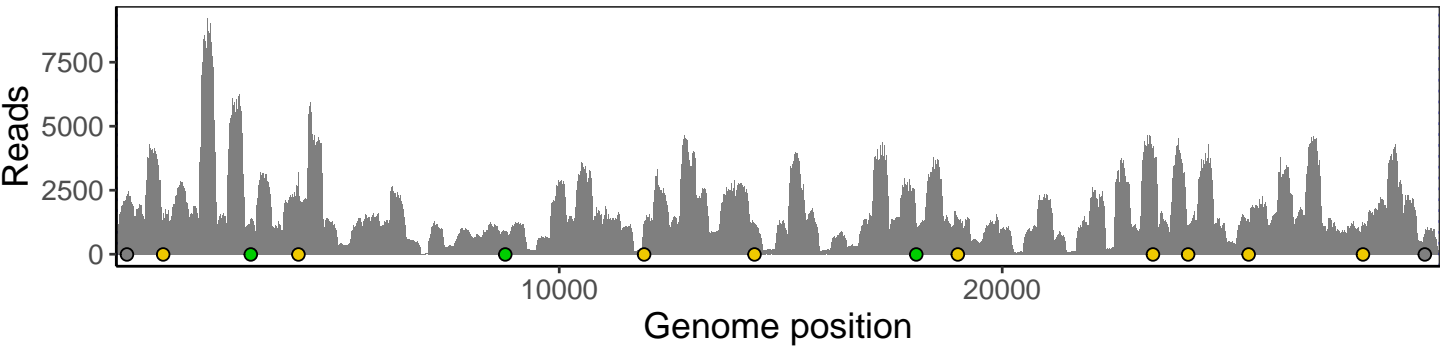

Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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

VSP0186-1m | 2020-04-13 | ETA | 7 | 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.

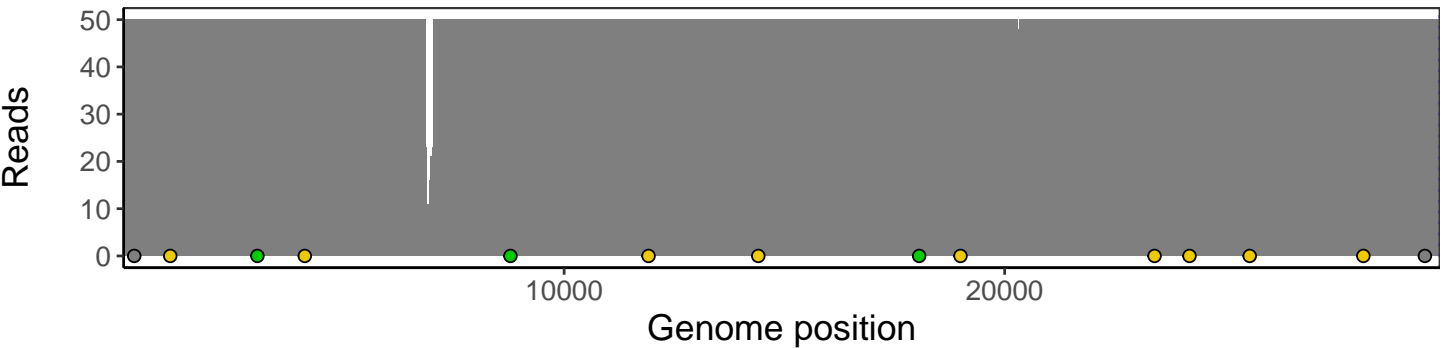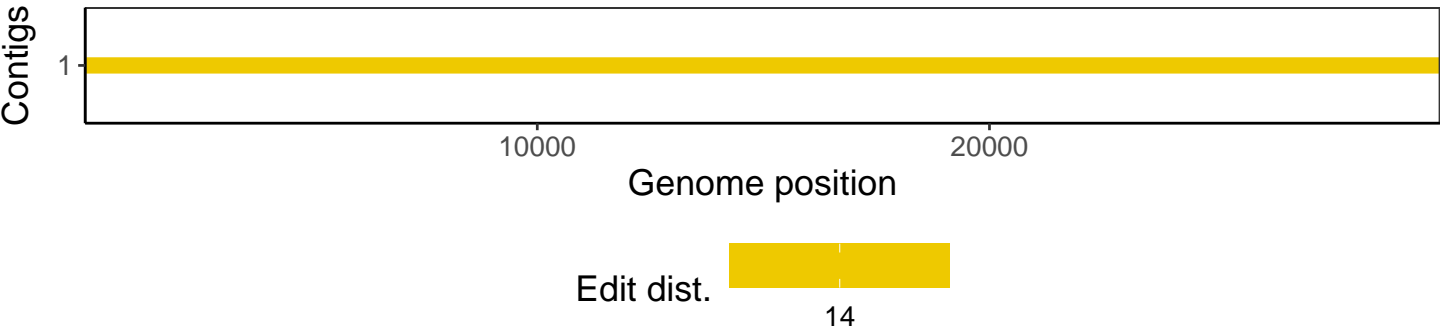Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

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