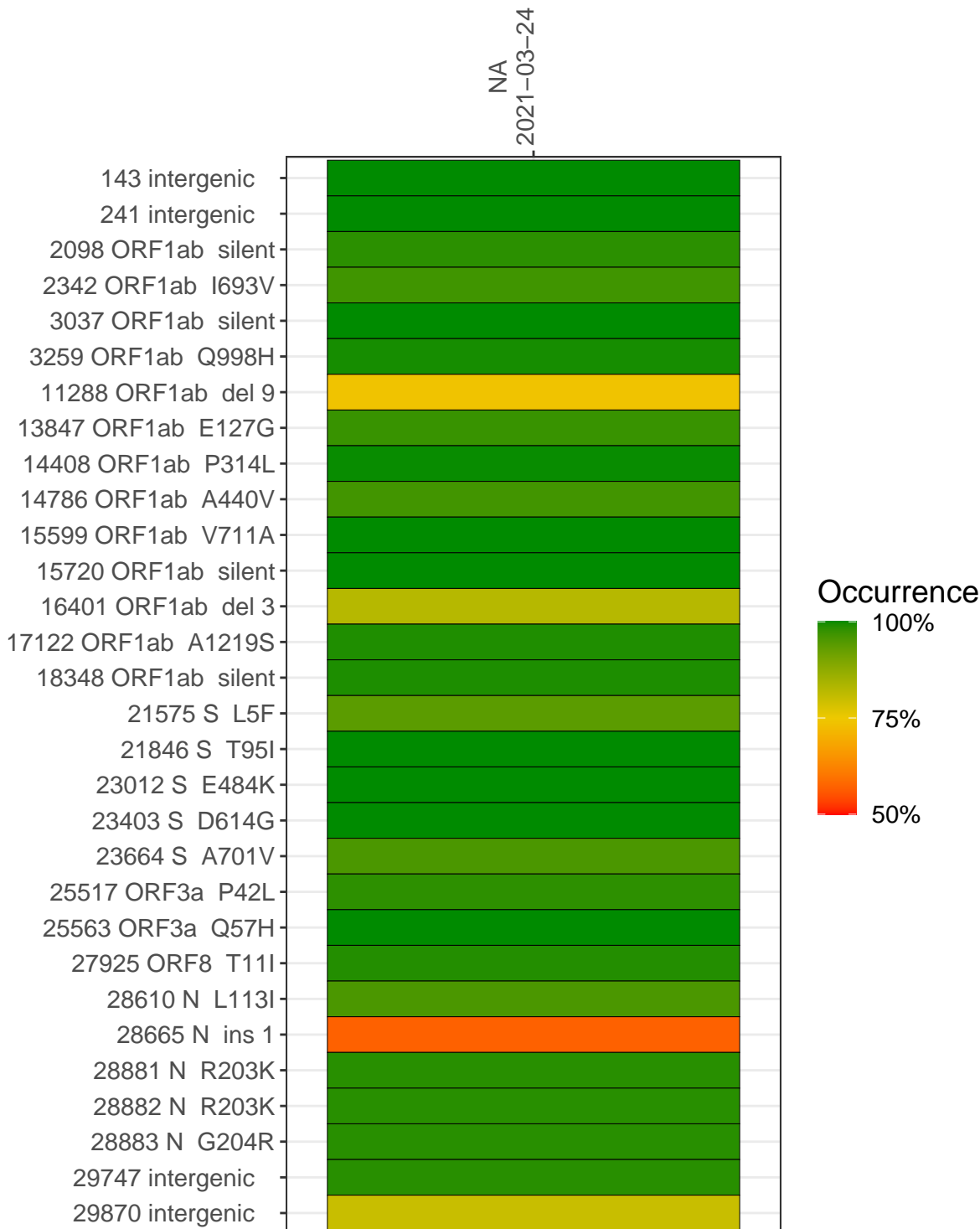# COVID-19 subject UPHS-0506

2021-06-23

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.
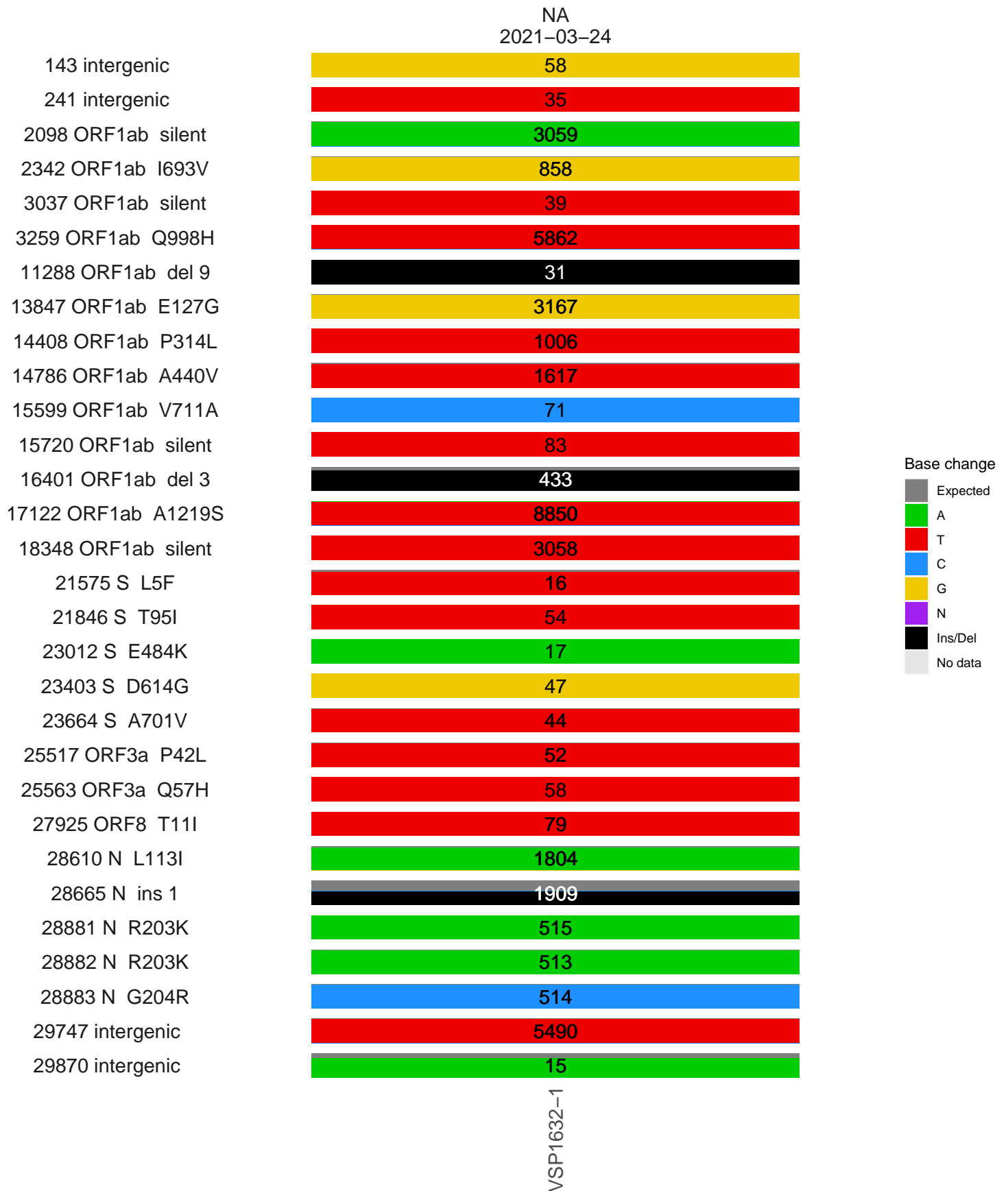
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1632-1	single experiment	NA	NA	2021-03-24	16.20	B.1.1.434	99.2%	98.8%

Variants shared across samples

The heat map below shows how variants (reference genome /home/common/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.

Position / Gene / Mutation	Value	Base change
143 intergenic	58	G
241 intergenic	35	T
2098 ORF1ab silent	3059	A
2342 ORF1ab I693V	858	G
3037 ORF1ab silent	39	T
3259 ORF1ab Q998H	5862	T
11288 ORF1ab del 9	31	Ins/Del
13847 ORF1ab E127G	3167	G
14408 ORF1ab P314L	1006	T
14786 ORF1ab A440V	1617	T
15599 ORF1ab V711A	71	C
15720 ORF1ab silent	83	T
16401 ORF1ab del 3	433	Ins/Del
17122 ORF1ab A1219S	8850	T
18348 ORF1ab silent	3058	T
21575 S L5F	16	T
21846 S T95I	54	T
23012 S E484K	17	A
23403 S D614G	47	G
23664 S A701V	44	T
25517 ORF3a P42L	52	T
25563 ORF3a Q57H	58	T
27925 ORF8 T11I	79	T
28610 N L113I	1804	A
28665 N ins 1	1909	Ins/Del (Expected)
28881 N R203K	515	A
28882 N R203K	513	A
28883 N G204R	514	C
29747 intergenic	5490	T
29870 intergenic	15	A

Base change

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

VSP1632−1

Analyses of individual experiments and composite results

VSP1632-1 | 2021-03-24 | NA | UPHS-0506 | 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.
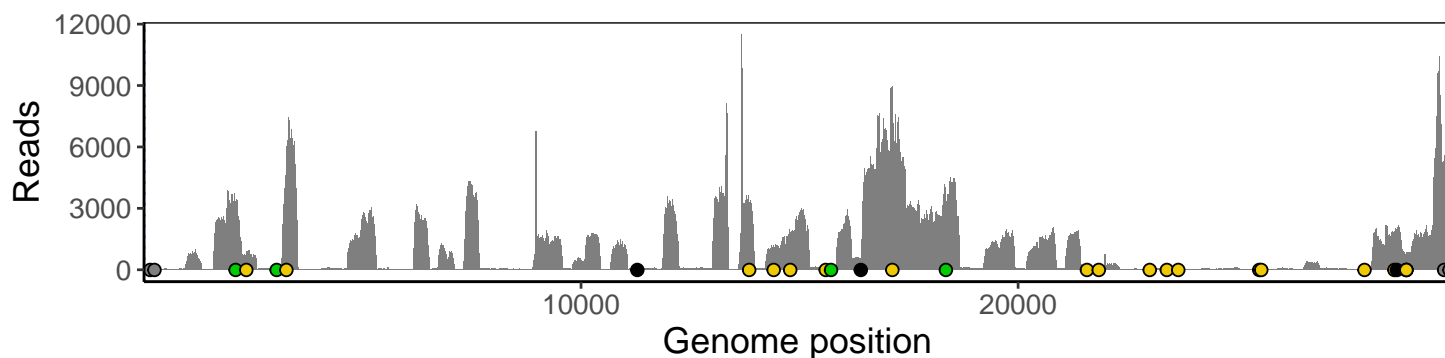
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
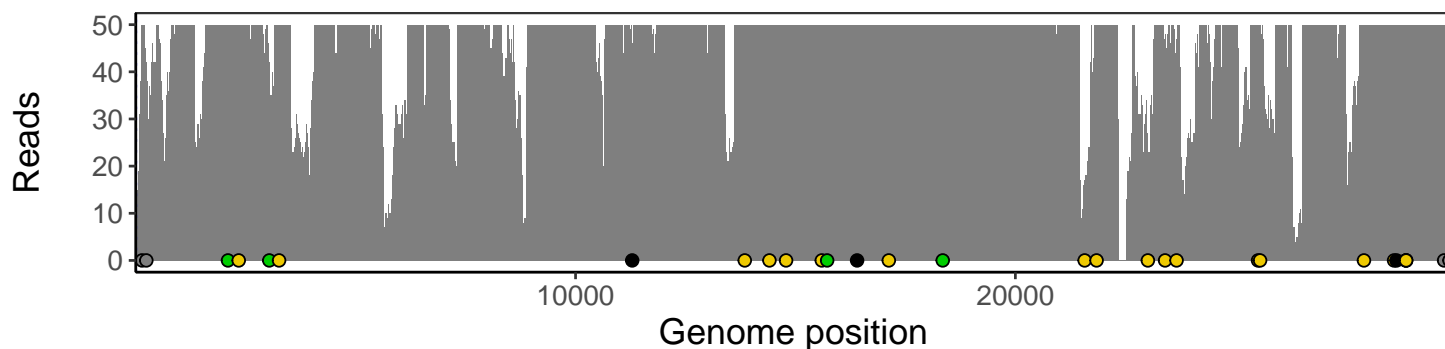
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
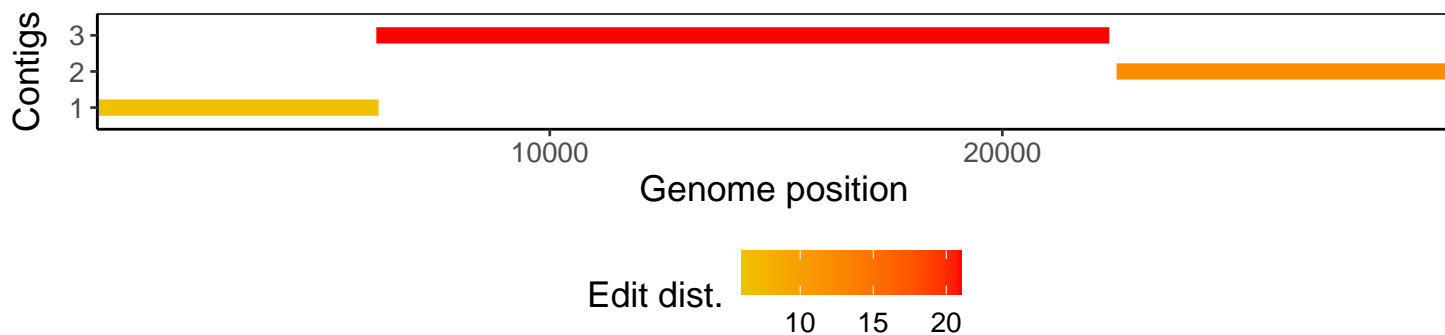
Software environment

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	3.1.3
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	3.3.3
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