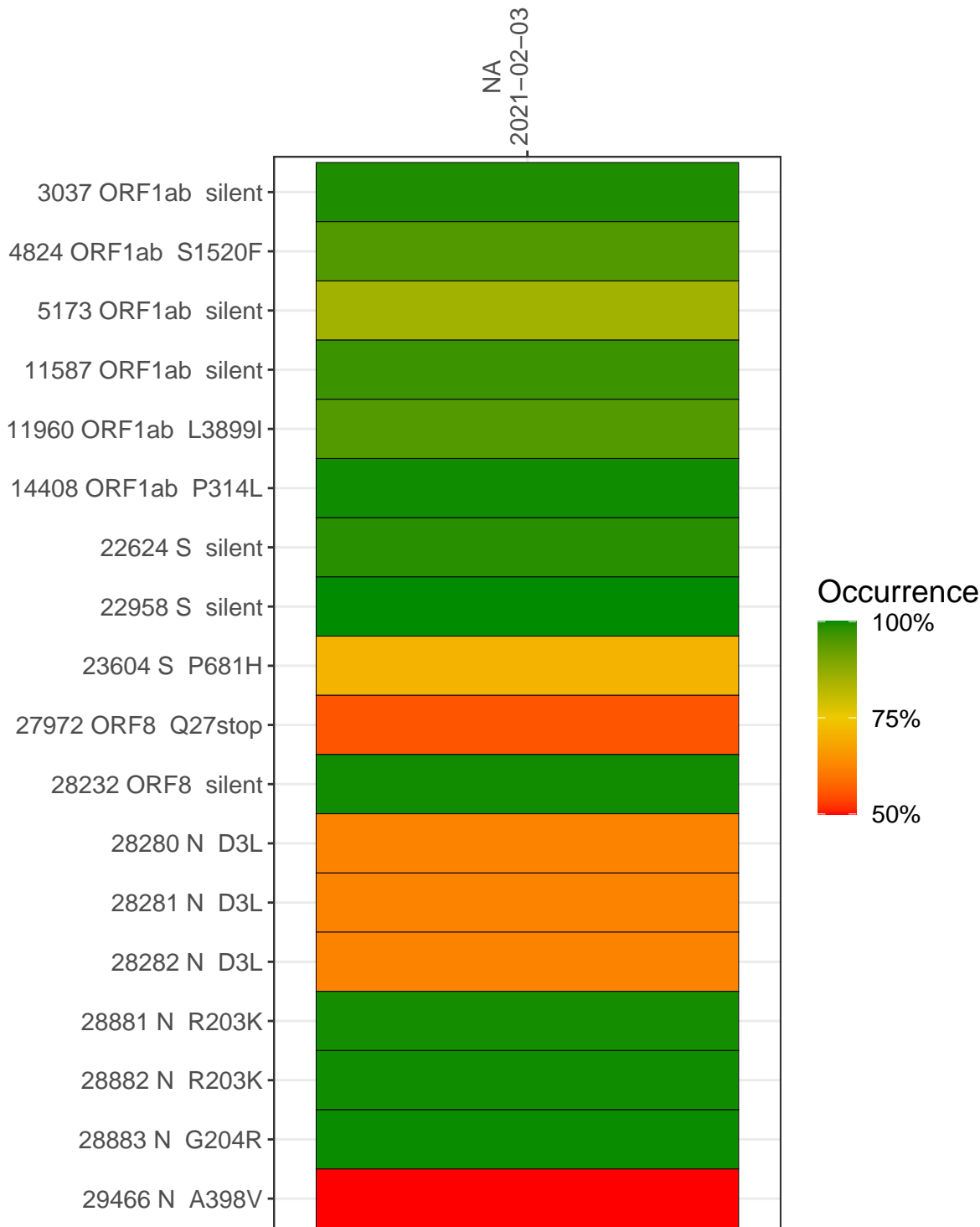# COVID-19 subject SARS_CoV_96

2021-06-29

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.
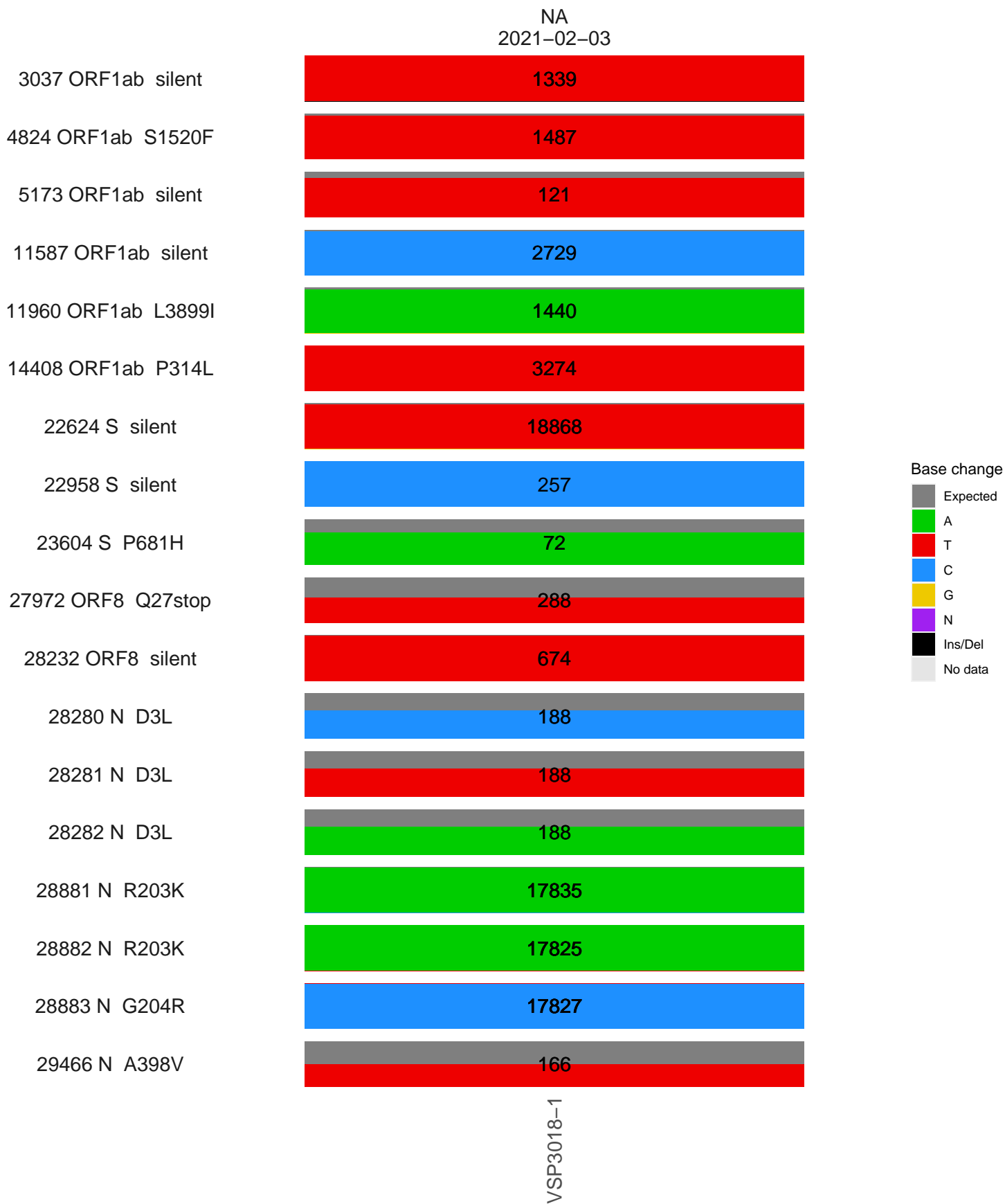
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP3018-1	single experiment	NA	NA	2021-02-03	2.62	B.1.1.161	99.7%	99.3%

Variants shared across samples

The heat map below shows how variants (reference genome /home/common/SARS-CoV-2-Philadelphia/NC_04551
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.

NA
2021−02−03

Position	Count
3037 ORF1ab silent	1339
4824 ORF1ab S1520F	1487
5173 ORF1ab silent	121
11587 ORF1ab silent	2729
11960 ORF1ab L3899I	1440
14408 ORF1ab P314L	3274
22624 S silent	18868
22958 S silent	257
23604 S P681H	72
27972 ORF8 Q27stop	288
28232 ORF8 silent	674
28280 N D3L	188
28281 N D3L	188
28282 N D3L	188
28881 N R203K	17835
28882 N R203K	17825
28883 N G204R	17827
29466 N A398V	166

VSP3018−1

Base change

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

3

Analyses of individual experiments and composite results

**VSP3018-1 | 2021-02-03 | NA | SARS_CoV_96 | 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.
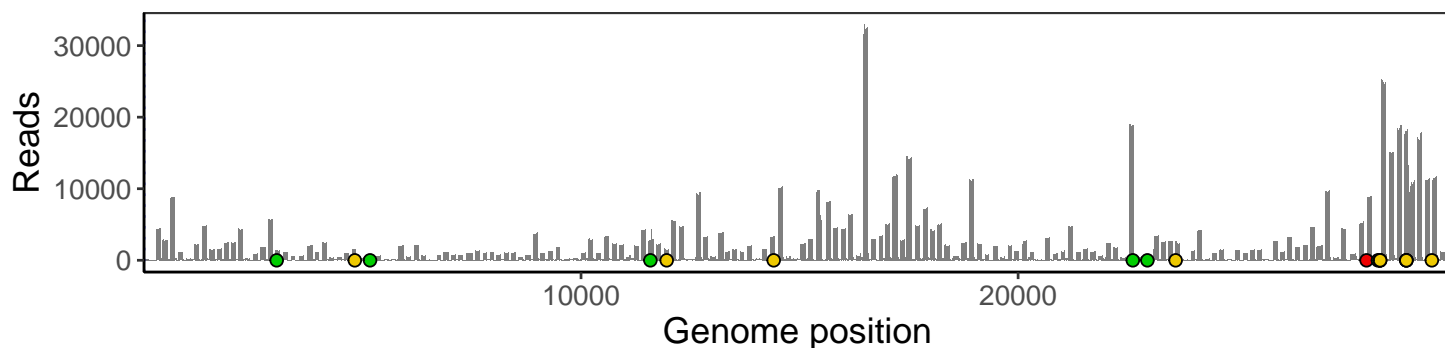
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
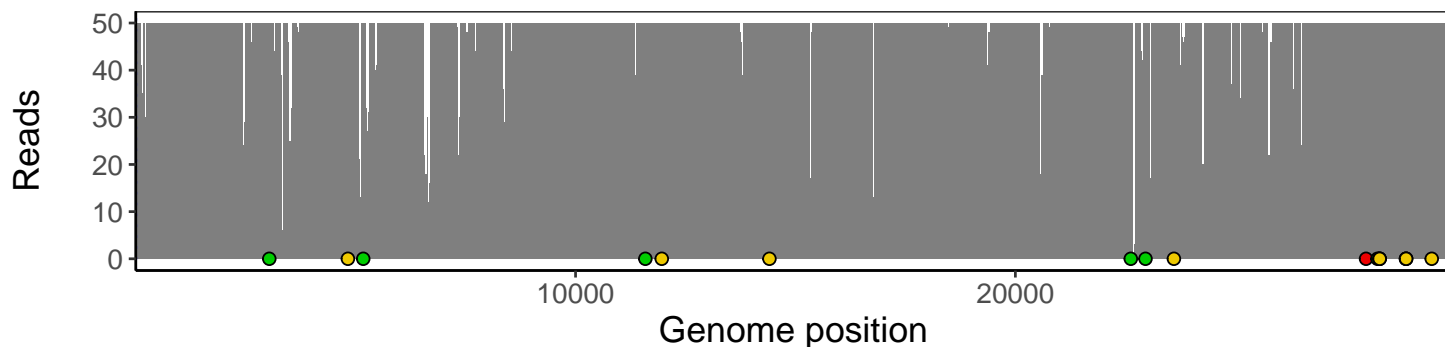
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.
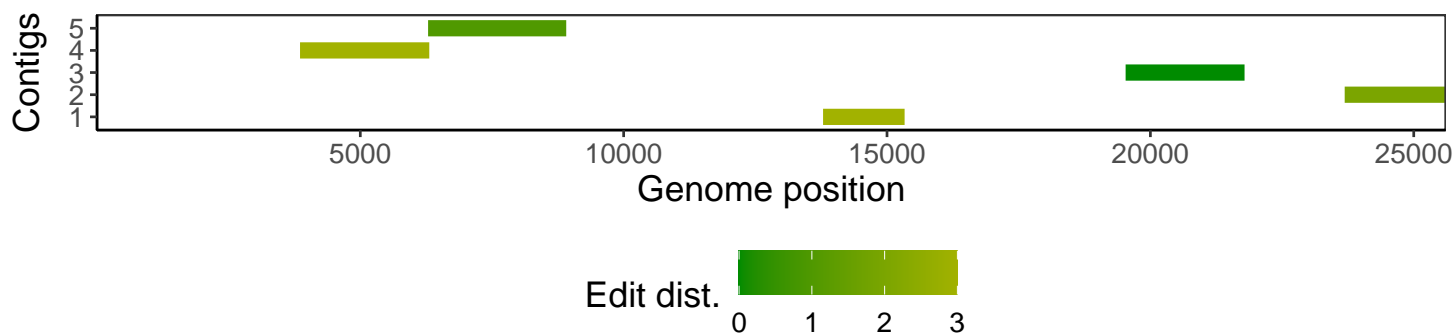
Software environment

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