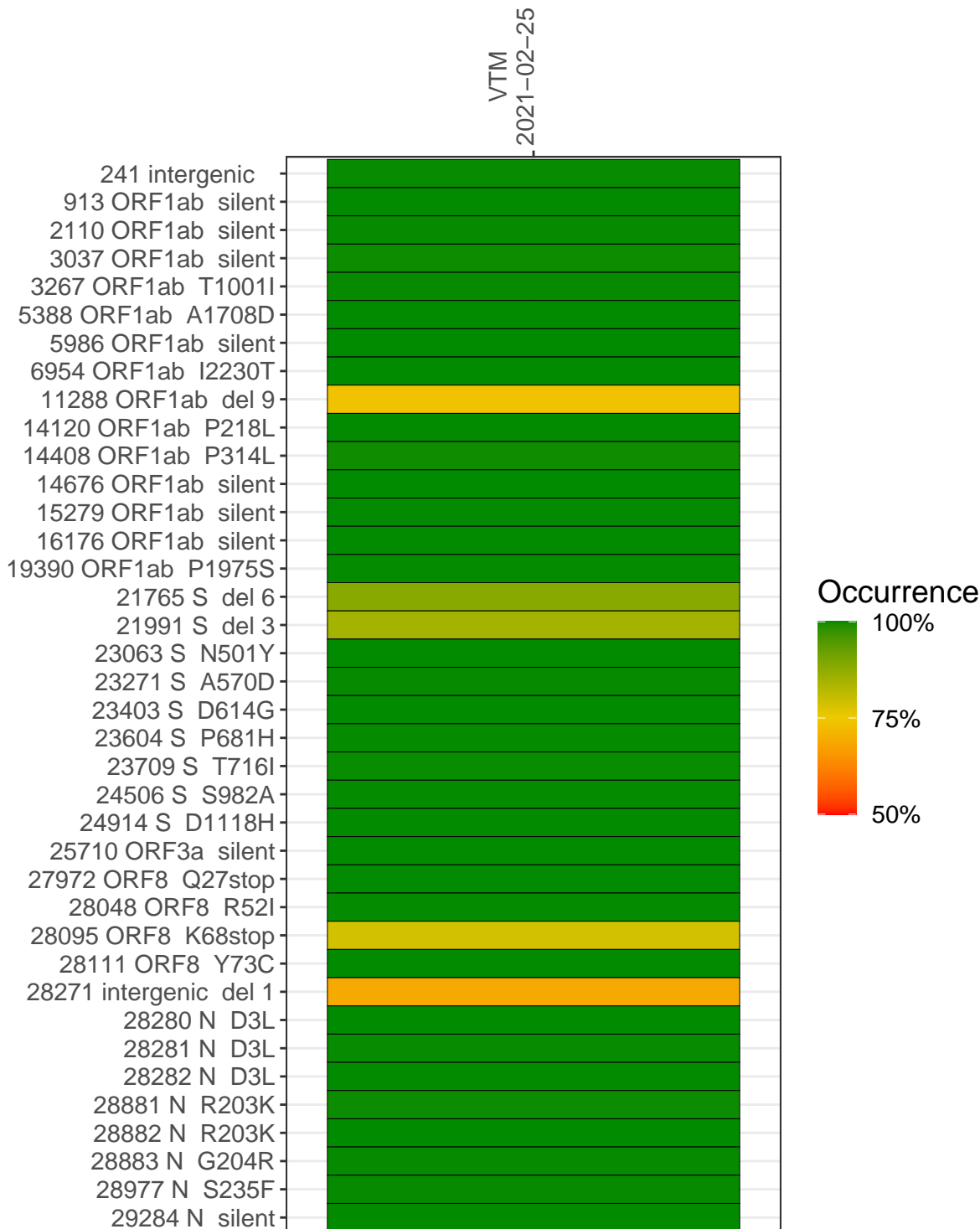# COVID-19 subject HUP Q-0024

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.
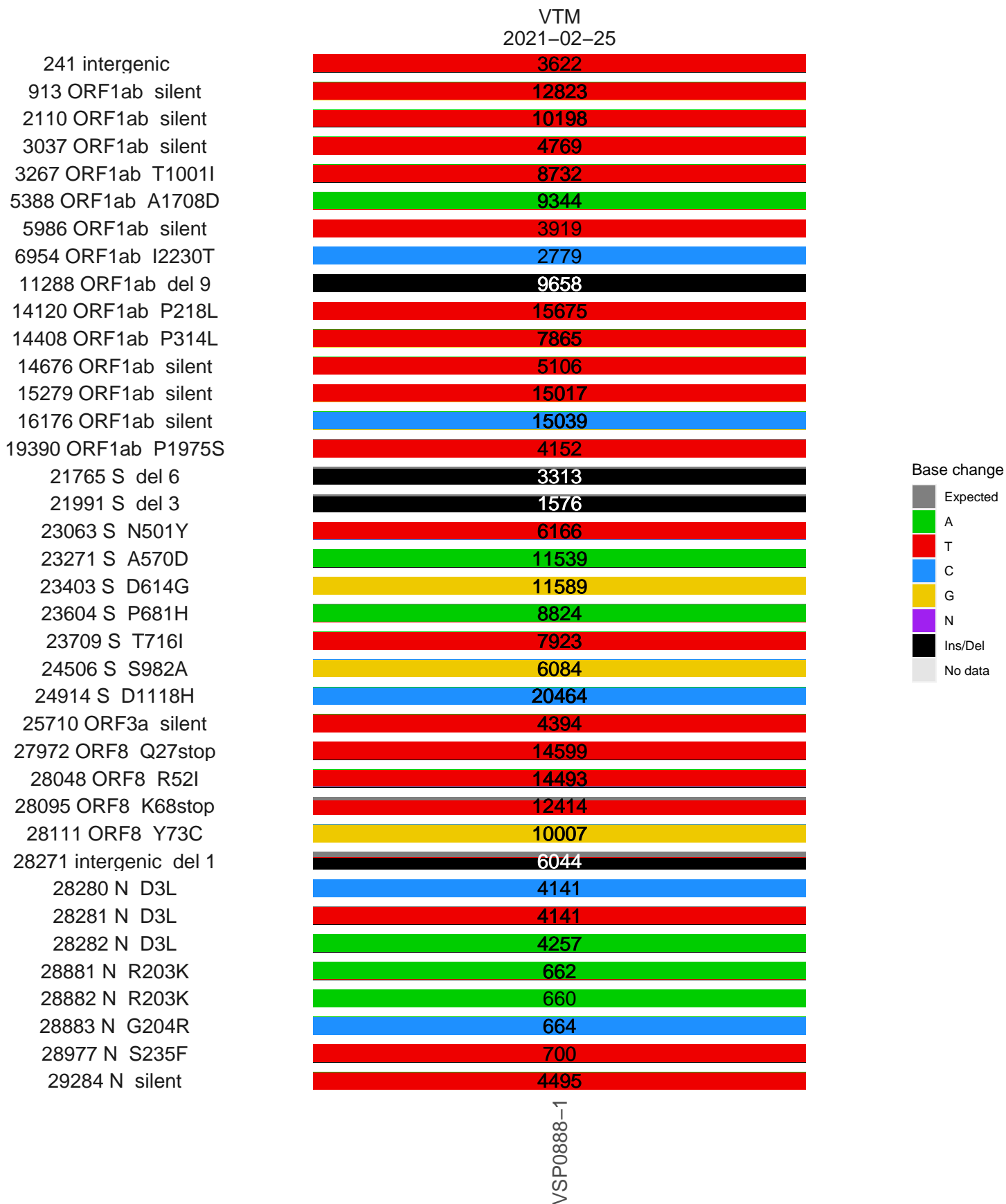
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0888-1	single experiment	NA	VTM	2021-02-25	29.90	B.1.1.7	100.0%	100.0%

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.

VTM
2021−02−25

Position / Gene / Change	Value	Base change
241 intergenic	3622	T
913 ORF1ab silent	12823	T
2110 ORF1ab silent	10198	A
3037 ORF1ab silent	4769	T
3267 ORF1ab T1001I	8732	T
5388 ORF1ab A1708D	9344	A
5986 ORF1ab silent	3919	T
6954 ORF1ab I2230T	2779	C
11288 ORF1ab del 9	9658	Ins/Del
14120 ORF1ab P218L	15675	T
14408 ORF1ab P314L	7865	T
14676 ORF1ab silent	5106	T
15279 ORF1ab silent	15017	T
16176 ORF1ab silent	15039	C
19390 ORF1ab P1975S	4152	T
21765 S del 6	3313	Ins/Del
21991 S del 3	1576	Ins/Del
23063 S N501Y	6166	T
23271 S A570D	11539	A
23403 S D614G	11589	G
23604 S P681H	8824	A
23709 S T716I	7923	T
24506 S S982A	6084	G
24914 S D1118H	20464	C
25710 ORF3a silent	4394	T
27972 ORF8 Q27stop	14599	T
28048 ORF8 R52I	14493	T
28095 ORF8 K68stop	12414	T
28111 ORF8 Y73C	10007	G
28271 intergenic del 1	6044	Ins/Del
28280 N D3L	4141	C
28281 N D3L	4141	T
28282 N D3L	4257	A
28881 N R203K	662	A
28882 N R203K	660	A
28883 N G204R	664	C
28977 N S235F	700	T
29284 N silent	4495	T

Base change

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

VSP0888−1

3

Analyses of individual experiments and composite results

VSP0888-1 | 2021-02-25 | VTM | HUP Q-0024 | 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.
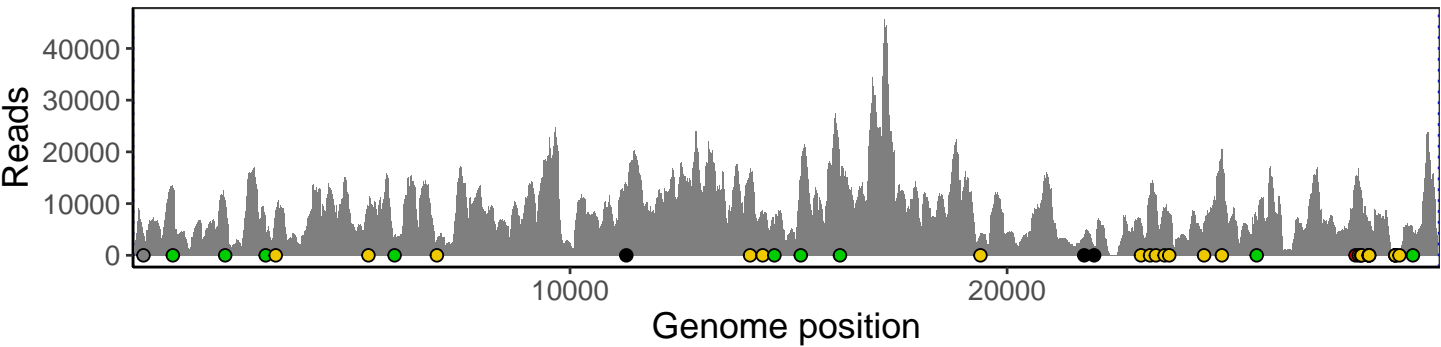
Excerpt from plot above focusing on reads coverage from 0 to 50 NT.
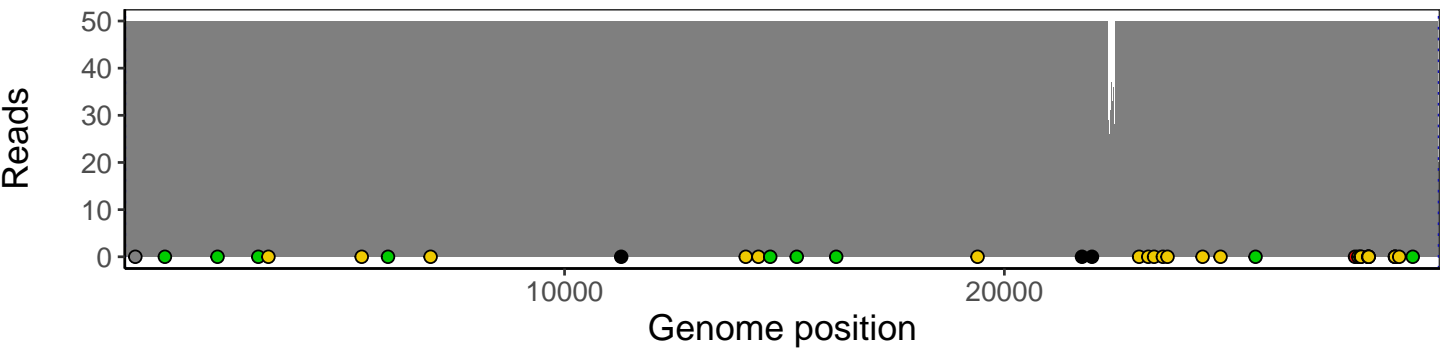
The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

4
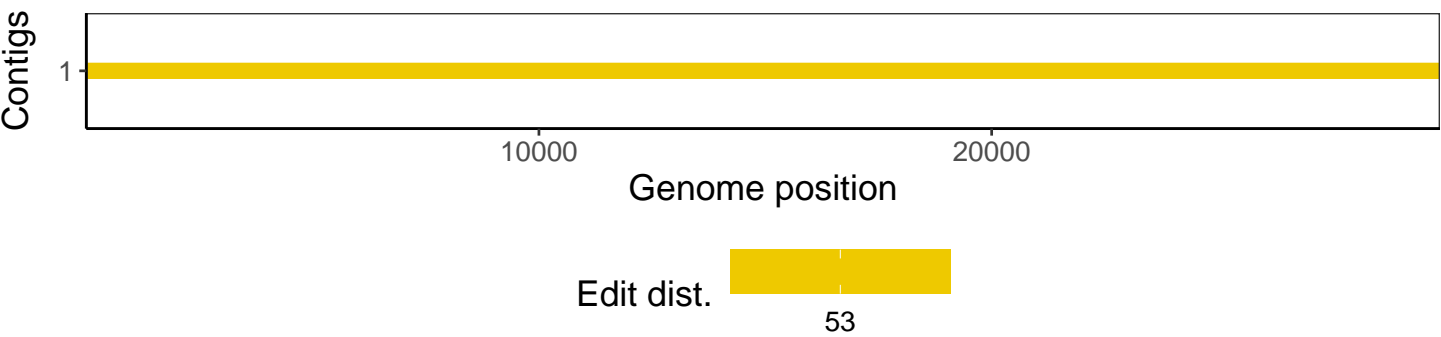
Software environment

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