COVID-19 subject UPHS-0160

2021-06-23

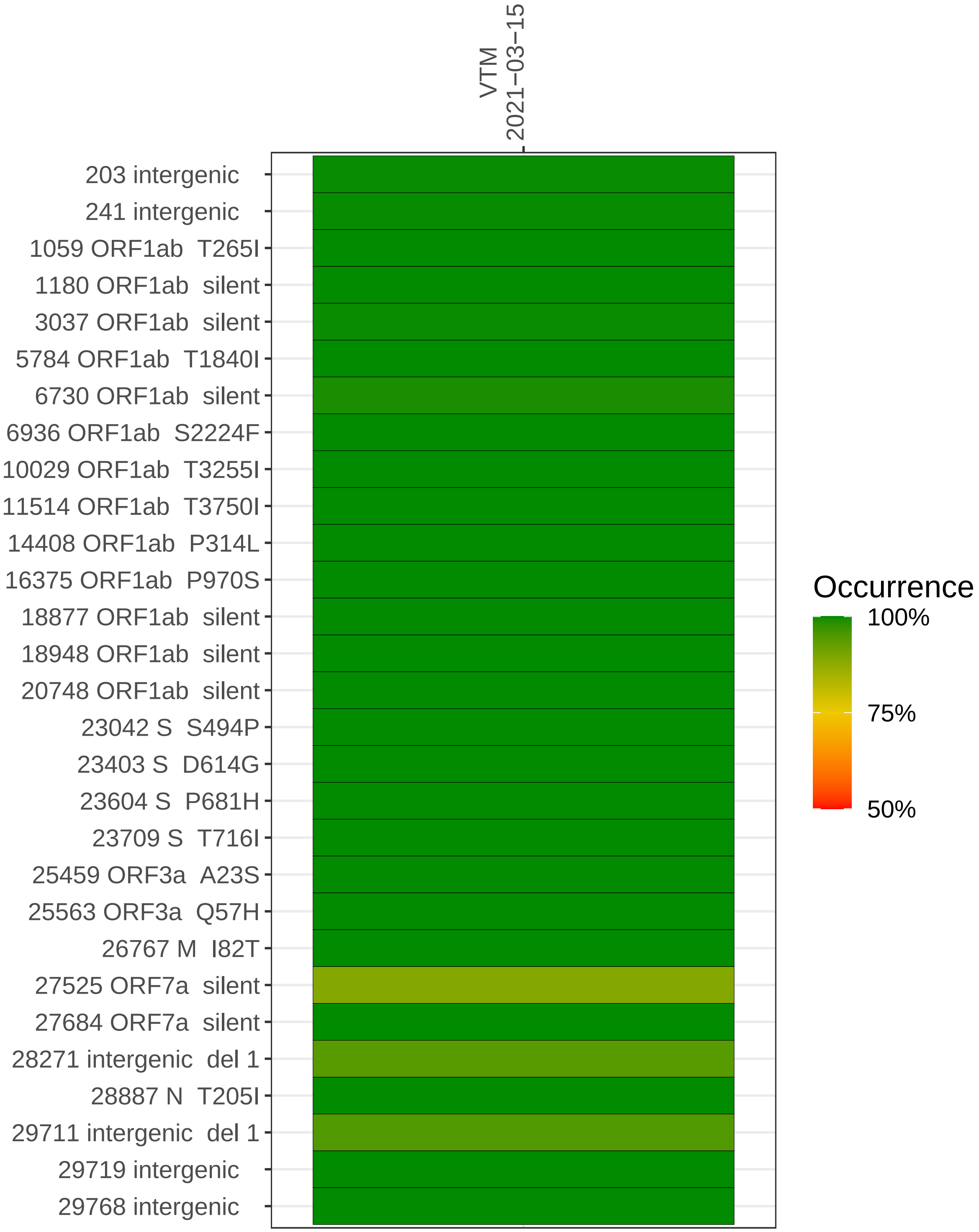
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software tool (Rambaut et al 2020) for genomes with > 90% sequence coverage.

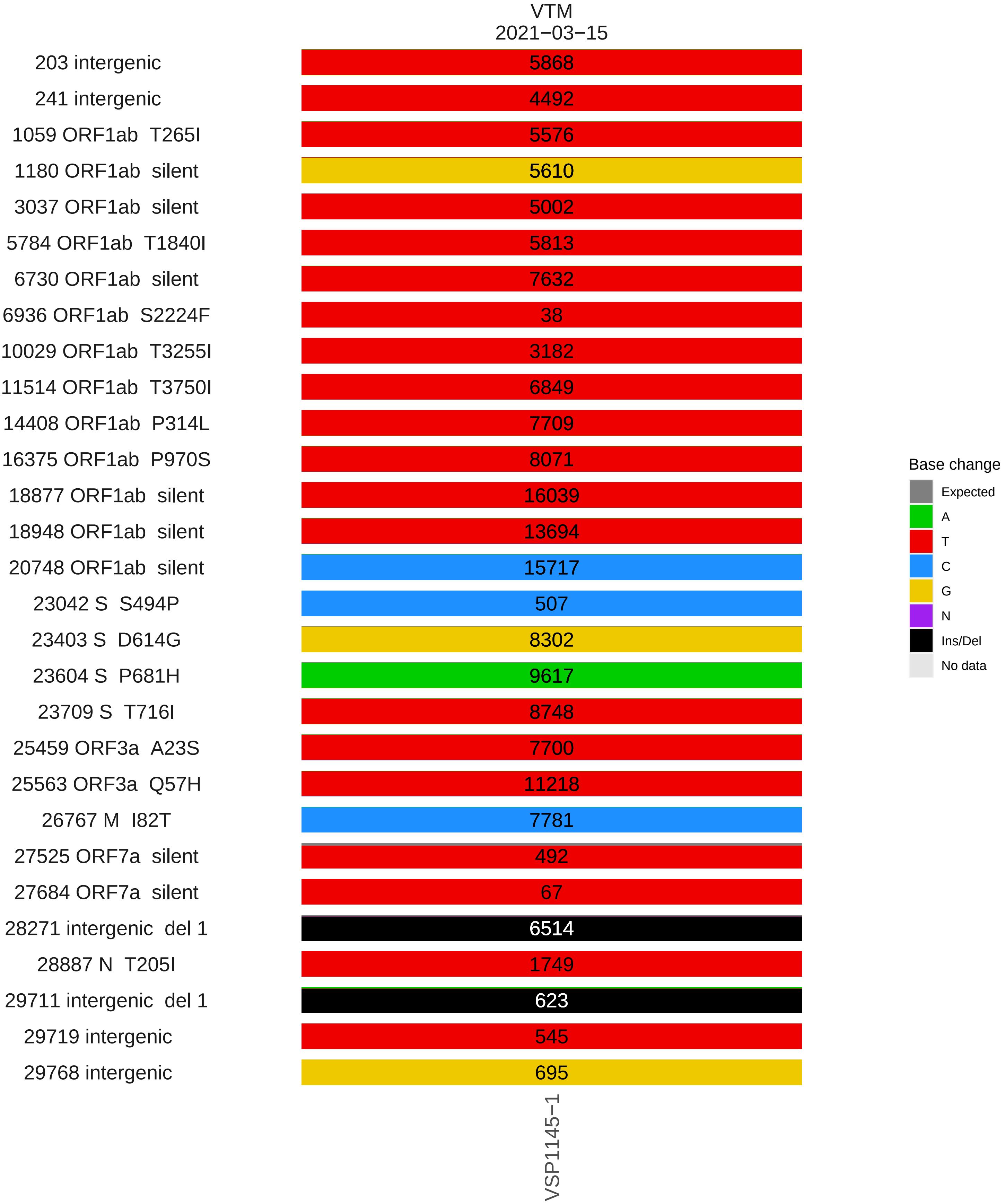
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1145-1	single experiment	NA	VTM	2021-03-15	29.68	B.1.575	99.3%	99.2%

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.

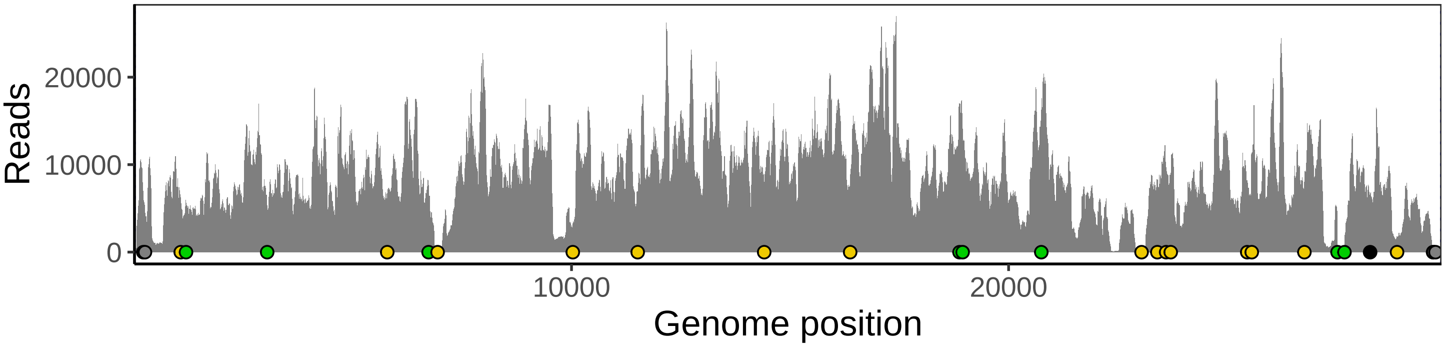



VTM
2021−03−15
203 intergenic 5868
241 intergenic 4492
1059 ORF1ab T265I 5576
1180 ORF1ab silent 5610
3037 ORF1ab silent 5002
5784 ORF1ab T1840I 5813
6730 ORF1ab silent 7632
6936 ORF1ab S2224F 38
10029 ORF1ab T3255I 3182
11514 ORF1ab T3750I 6849
14408 ORF1ab P314L 7709
16375 ORF1ab P970S 8071
18877 ORF1ab silent 16039
18948 ORF1ab silent 13694
20748 ORF1ab silent 15717
23042 S S494P 507
23403 S D614G 8302
23604 S P681H 9617
23709 S T716I 8748
25459 ORF3a A23S 7700
25563 ORF3a Q57H 11218
26767 M I82T 7781
27525 ORF7a silent 492
27684 ORF7a silent 67
28271 intergenic del 1 6514
28887 N T205I 1749
29711 intergenic del 1 623
29719 intergenic 545
29768 intergenic 695
VSP1145−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

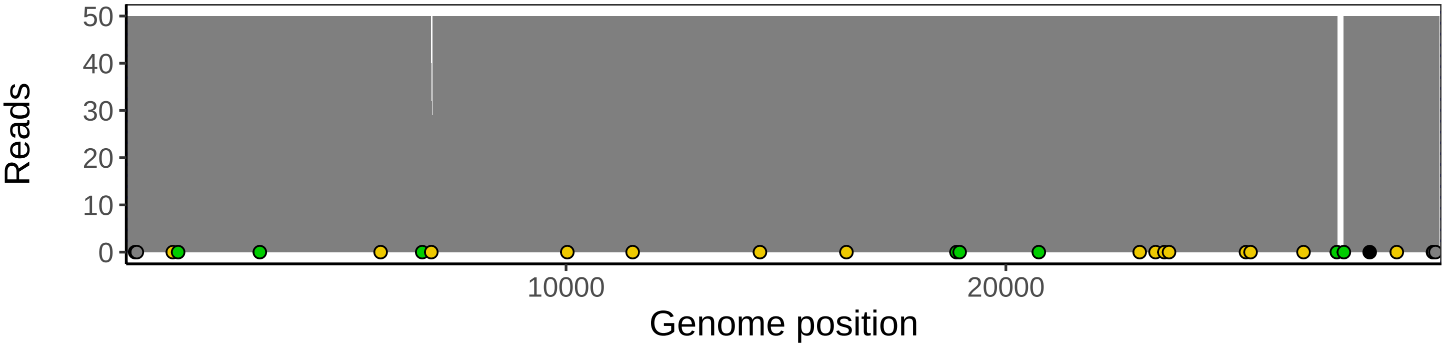
Analyses of individual experiments and composite results

VSP1145-1 | 2021-03-15 | VTM | UPHS-0160 | genomes | single experiment

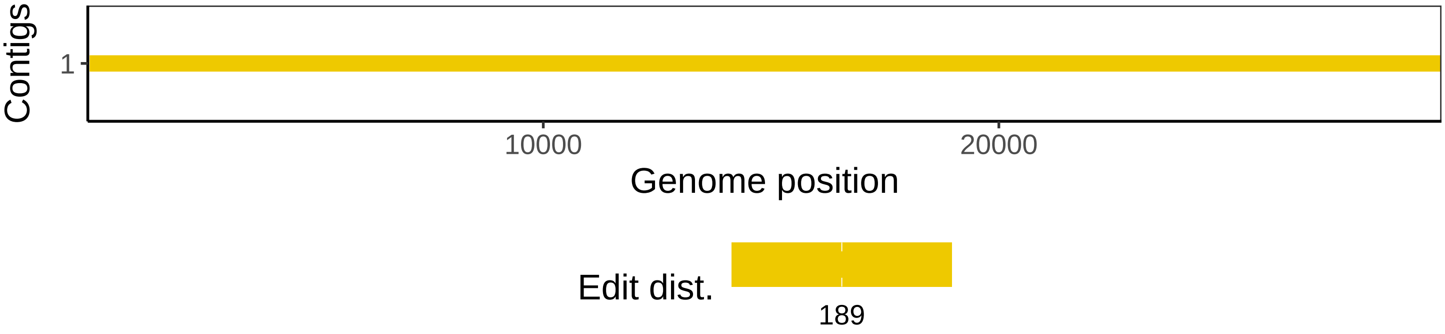
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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