COVID-19 subject H2101290600

2021-04-17

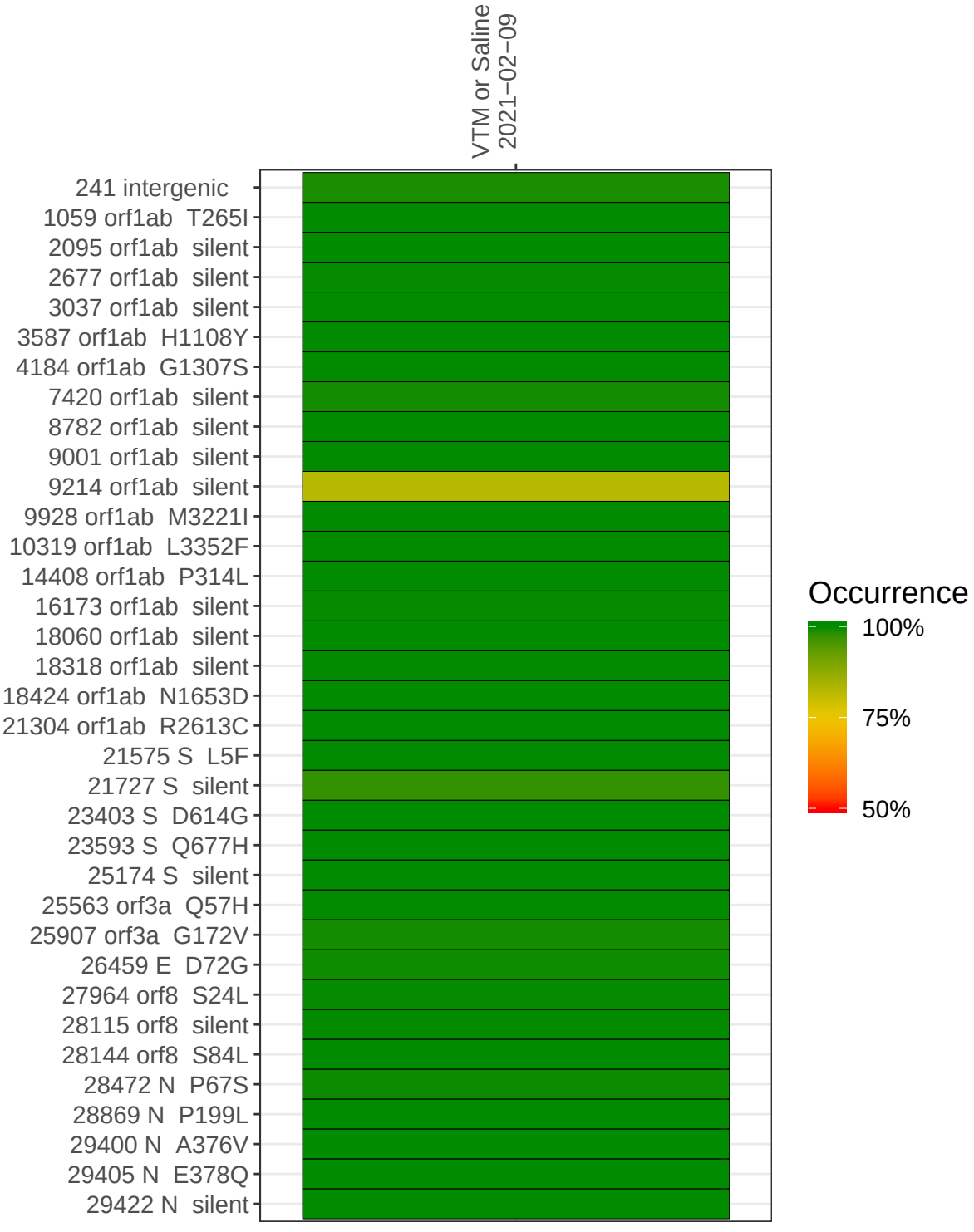
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

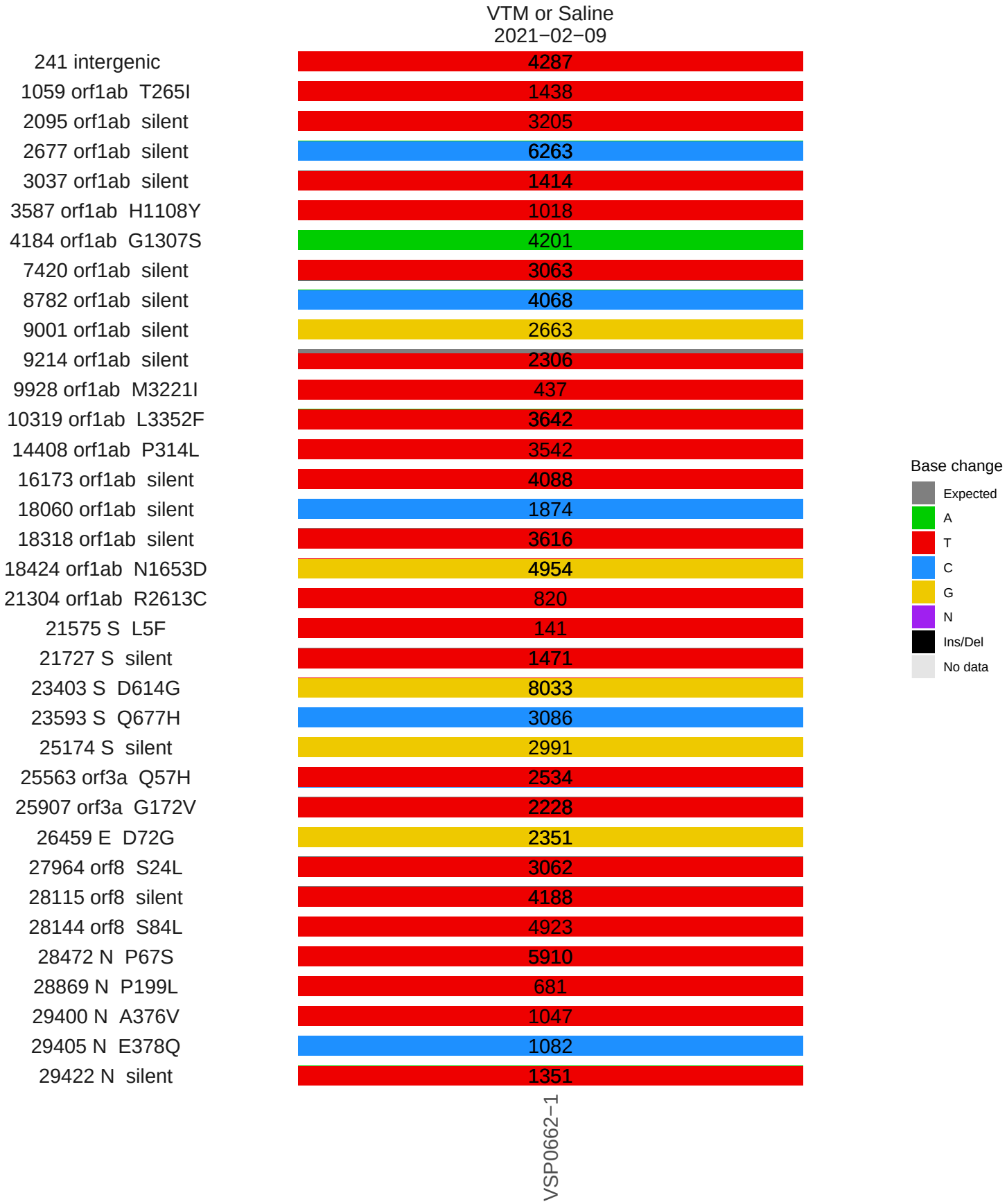
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0662-1	single experiment	NA	VTM or Saline	2021-02-09	29.74	B.1.2	99.8%	99.8%

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.

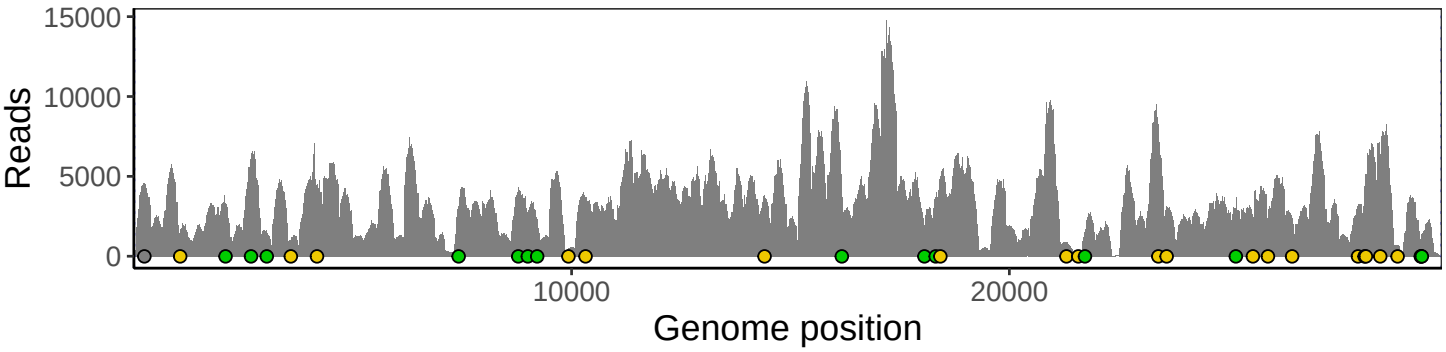



VTM or Saline
2021−02−09
241 intergenic
1059 orf1ab T265I
2095 orf1ab silent
2677 orf1ab silent
3037 orf1ab silent
3587 orf1ab H1108Y
4184 orf1ab G1307S
7420 orf1ab silent
8782 orf1ab silent
9001 orf1ab silent
9214 orf1ab silent
9928 orf1ab M3221I
10319 orf1ab L3352F
14408 orf1ab P314L
16173 orf1ab silent
18060 orf1ab silent
18318 orf1ab silent
18424 orf1ab N1653D
21304 orf1ab R2613C
21575 S L5F
21727 S silent
23403 S D614G
23593 S Q677H
25174 S silent
25563 orf3a Q57H
25907 orf3a G172V
26459 E D72G
27964 orf8 S24L
28115 orf8 silent
28144 orf8 S84L
28472 N P67S
28869 N P199L
29400 N A376V
29405 N E378Q
29422 N silent
4287
1438
3205
6263
1414
1018
4201
3063
4068
2663
2306
437
3642
3542
4088
1874
3616
4954
820
141
1471
8033
3086
2991
2534
2228
2351
3062
4188
4923
5910
681
1047
1082
1351
VSP0662−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

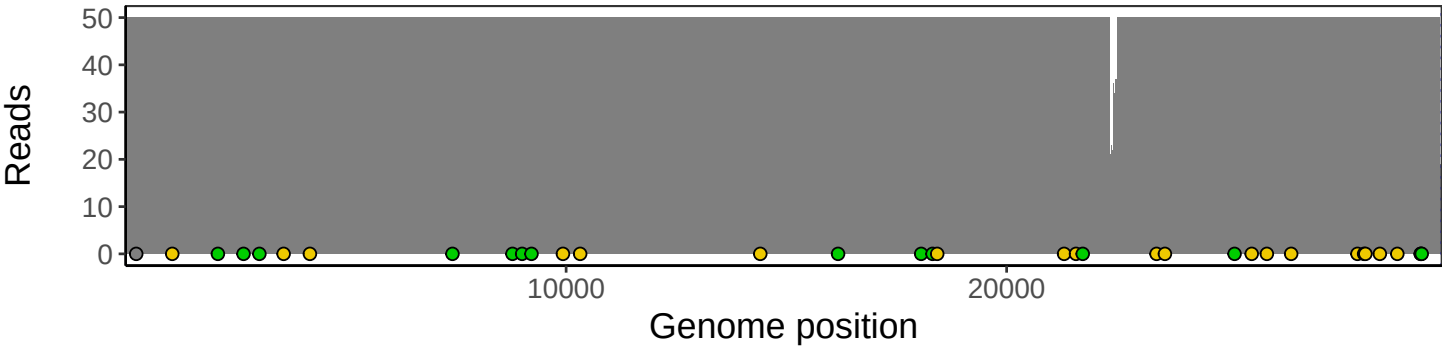
Analyses of individual experiments and composite results

VSP0662-1 | 2021-02-09 | VTM or Saline | H2101290600 | genomes | single experiment

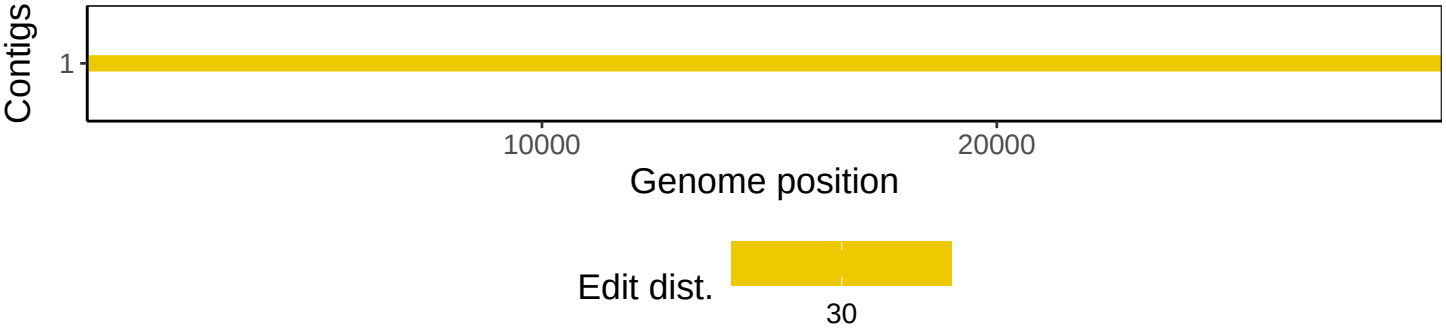
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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