COVID-19 subject H2101290600

2021-05-05

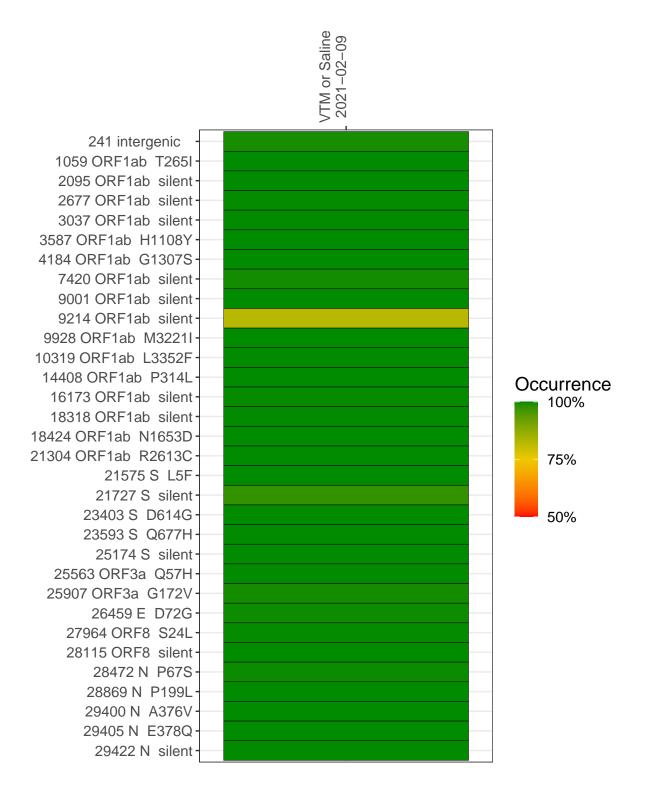
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

Table 1. Sample summary.

Experiment	Туре	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/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.



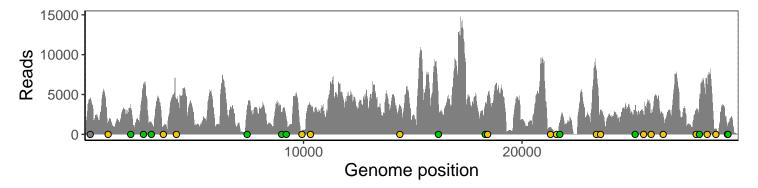
VTM or Saline 2021-02-09

	2021-02-09
241 intergenic	4287
1059 ORF1ab T265I	1438
2095 ORF1ab silent	3205
2677 ORF1ab silent	6263
3037 ORF1ab silent	1414
3587 ORF1ab H1108Y	1018
4184 ORF1ab G1307S	4201
7420 ORF1ab silent	3063
9001 ORF1ab silent	2663
9214 ORF1ab silent	2306
9928 ORF1ab M3221I	437
10319 ORF1ab L3352F	3642
14408 ORF1ab P314L	3542
16173 ORF1ab silent	4088
18318 ORF1ab silent	3616
18424 ORF1ab N1653D	4954
21304 ORF1ab R2613C	820
21575 S L5F	141
21727 S silent	1471
23403 S D614G	8033
23593 S Q677H	3086
25174 S silent	2991
25563 ORF3a Q57H	2534
25907 ORF3a G172V	2228
26459 E D72G	2351
27964 ORF8 S24L	3062
28115 ORF8 silent	4188
28472 N P67S	5910
28869 N P199L	681
29400 N A376V	1047
29405 N E378Q	1082
29422 N silent	1351
	2-1
	1.2°

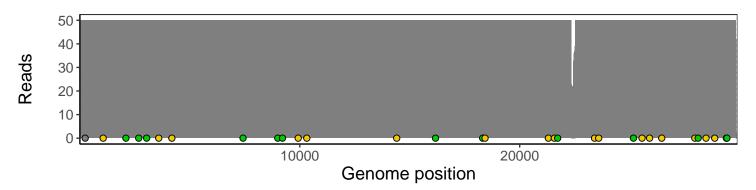
Analyses of individual experiments and composite results

$VSP0662\text{-}1 \mid 2021\text{-}02\text{-}09 \mid VTM \text{ or Saline} \mid H2101290600 \mid \text{genomes} \mid \text{single experiment}$

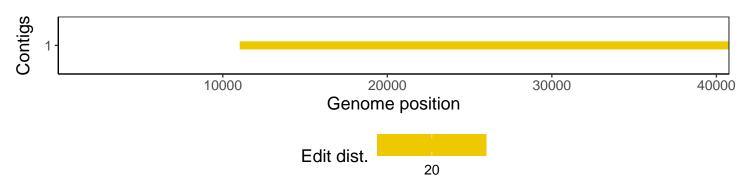
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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
${\it Genomic Alignments}$	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
$\operatorname{GenomeInfoDb}$	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