COVID-19 subject SARS_CoV_129

2021-06-29

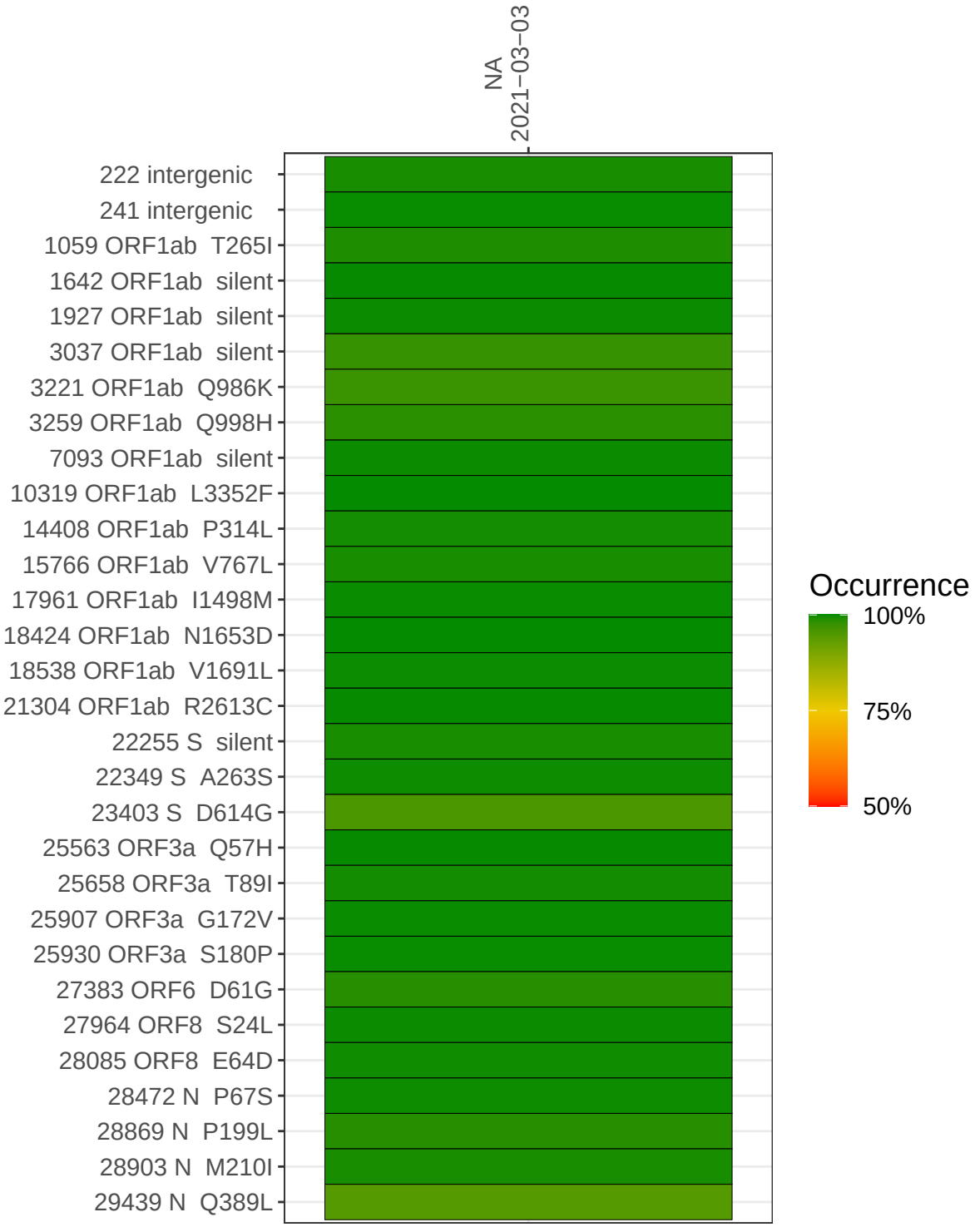
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

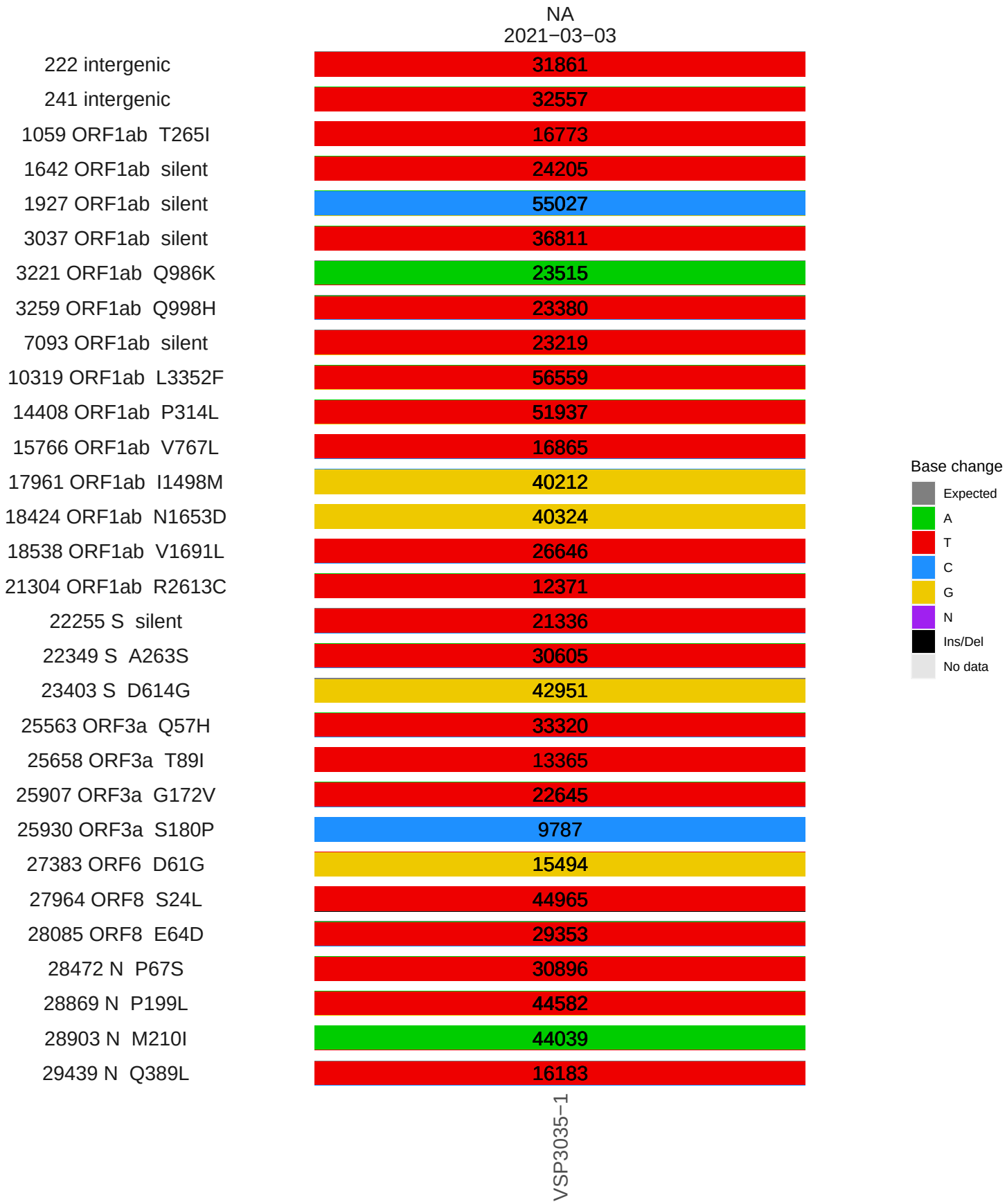
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP3035-1	single experiment	NA	NA	2021-03-03	28.07	B.1.2	99.9%	99.8%

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.

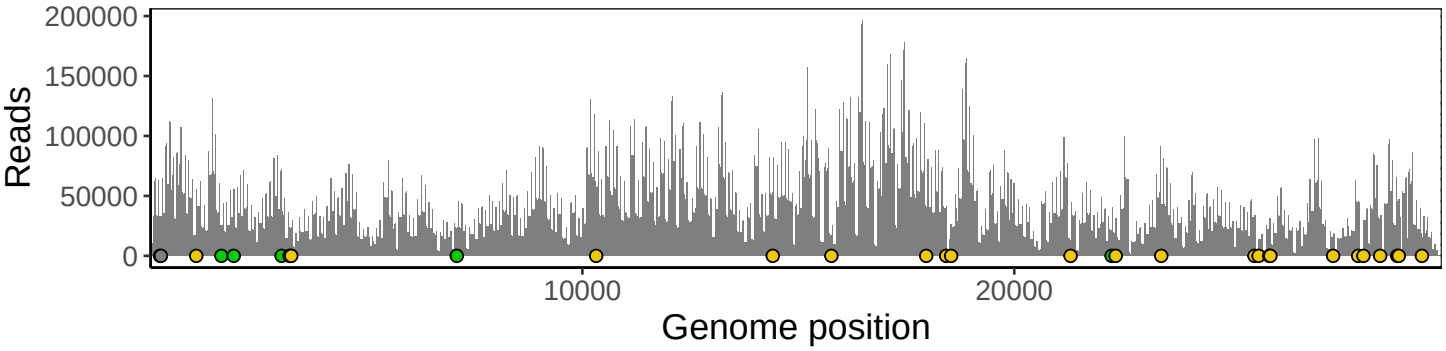



NA
2021−03−03
222 intergenic
241 intergenic
1059 ORF1ab T265I
1642 ORF1ab silent
1927 ORF1ab silent
3037 ORF1ab silent
3221 ORF1ab Q986K
3259 ORF1ab Q998H
7093 ORF1ab silent
10319 ORF1ab L3352F
14408 ORF1ab P314L
15766 ORF1ab V767L
17961 ORF1ab I1498M
18424 ORF1ab N1653D
18538 ORF1ab V1691L
21304 ORF1ab R2613C
22255 S silent
22349 S A263S
23403 S D614G
25563 ORF3a Q57H
25658 ORF3a T89I
25907 ORF3a G172V
25930 ORF3a S180P
27383 ORF6 D61G
27964 ORF8 S24L
28085 ORF8 E64D
28472 N P67S
28869 N P199L
28903 N M210I
29439 N Q389L
31861
32557
16773
24205
55027
36811
23515
23380
23219
56559
51937
16865
40212
40324
26646
12371
21336
30605
42951
33320
13365
22645
9787
15494
44965
29353
30896
44582
44039
16183
VSP3035−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

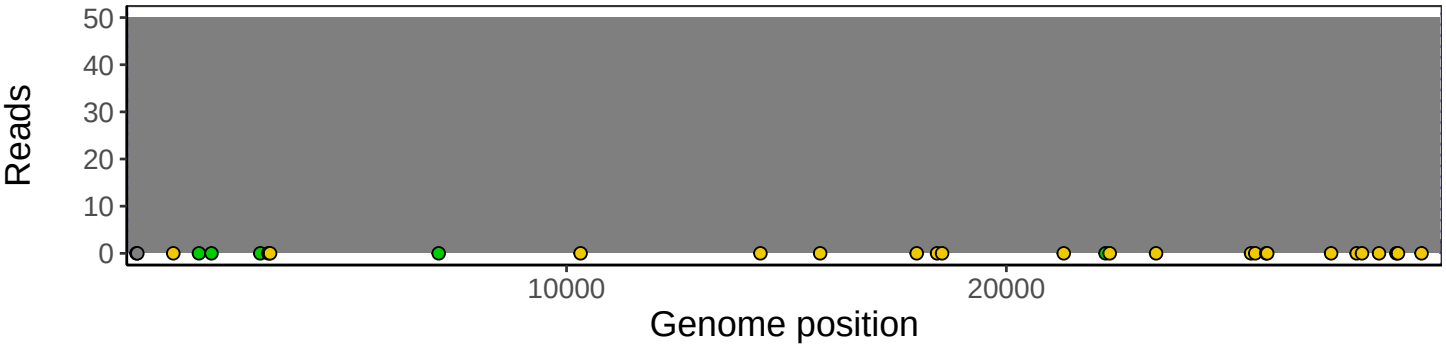
Analyses of individual experiments and composite results

VSP3035-1 | 2021-03-03 | NA | SARS_CoV_129 | genomes | single experiment

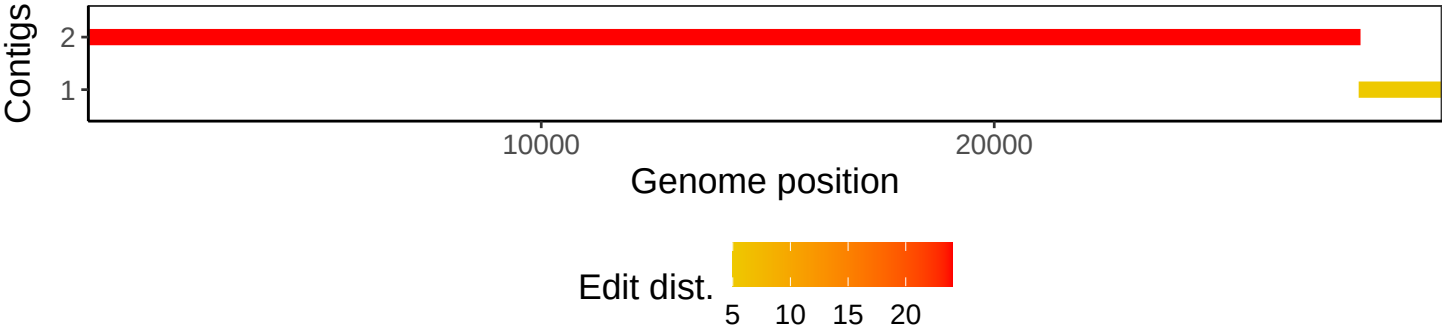
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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