COVID-19 subject SARS_CoV_262

2021-06-29

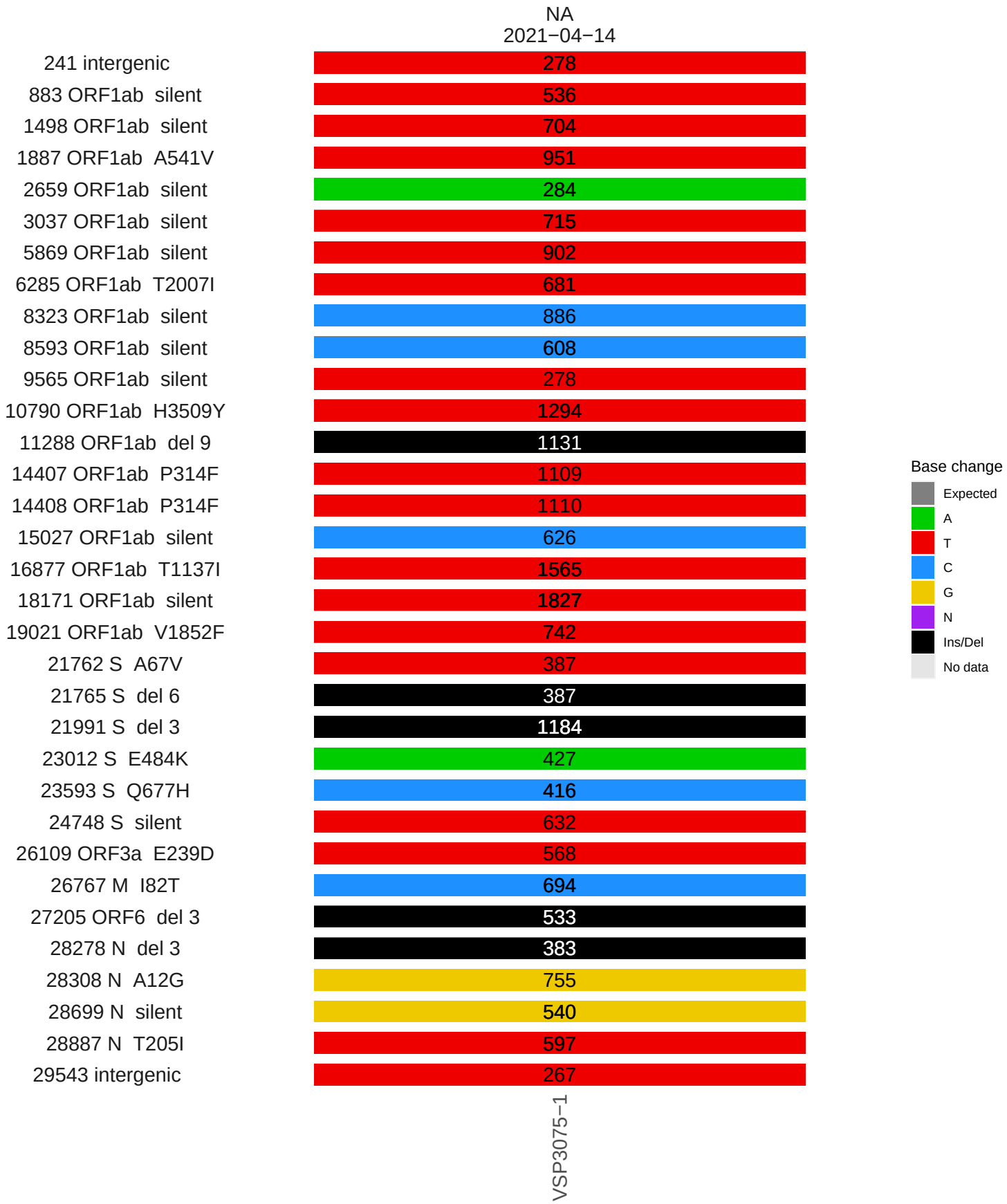
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP3075-1	single experiment	NA	NA	2021-04-14	29.76	B.1.525	99.6%	99.5%

Variants shared across samples

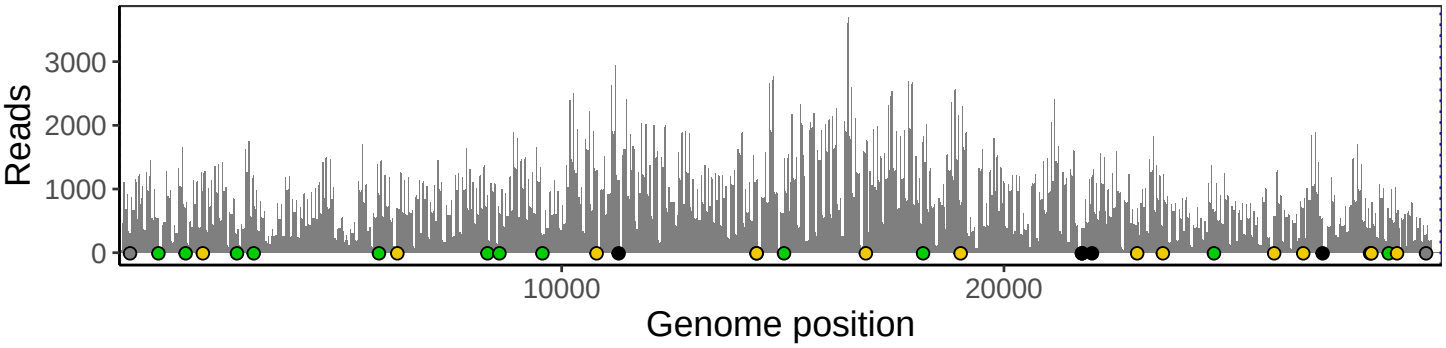
The heat map below shows how variants (reference genome /home/common/SARS-CoV-2-Philadelphia/NC_04551 are shared across subject samples where the percent variance is colored. Variants are called if a variant position is covered by 5 or more reads, the alternative base is found in $> 50\%$ of read pairs 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
2021−04−14
241 intergenic
278
883 ORF1ab silent
536
1498 ORF1ab silent
704
1887 ORF1ab A541V
951
2659 ORF1ab silent
284
3037 ORF1ab silent
715
5869 ORF1ab silent
902
6285 ORF1ab T2007I
681
8323 ORF1ab silent
886
8593 ORF1ab silent
608
9565 ORF1ab silent
278
10790 ORF1ab H3509Y
1294
11288 ORF1ab del 9
1131
14407 ORF1ab P314F
1109
14408 ORF1ab P314F
1110
15027 ORF1ab silent
626
16877 ORF1ab T1137I
1565
18171 ORF1ab silent
1827
19021 ORF1ab V1852F
742
21762 S A67V
387
21765 S del 6
387
21991 S del 3
1184
23012 S E484K
427
23593 S Q677H
416
24748 S silent
632
26109 ORF3a E239D
568
26767 M I82T
694
27205 ORF6 del 3
533
28278 N del 3
383
28308 N A12G
755
28699 N silent
540
28887 N T205I
597
29543 intergenic
267
VSP3075−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

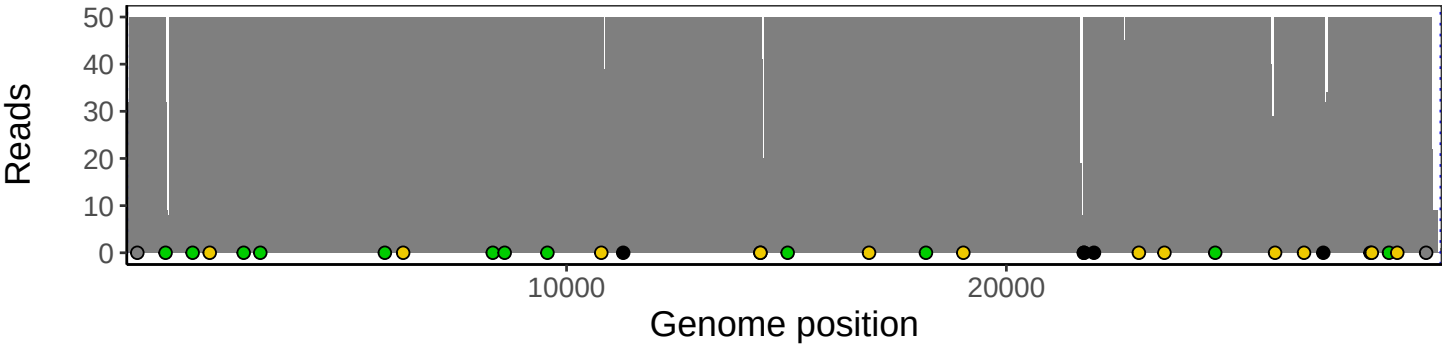
Analyses of individual experiments and composite results

VSP3075-1 | 2021-04-14 | NA | SARS_CoV_262 | genomes | single experiment

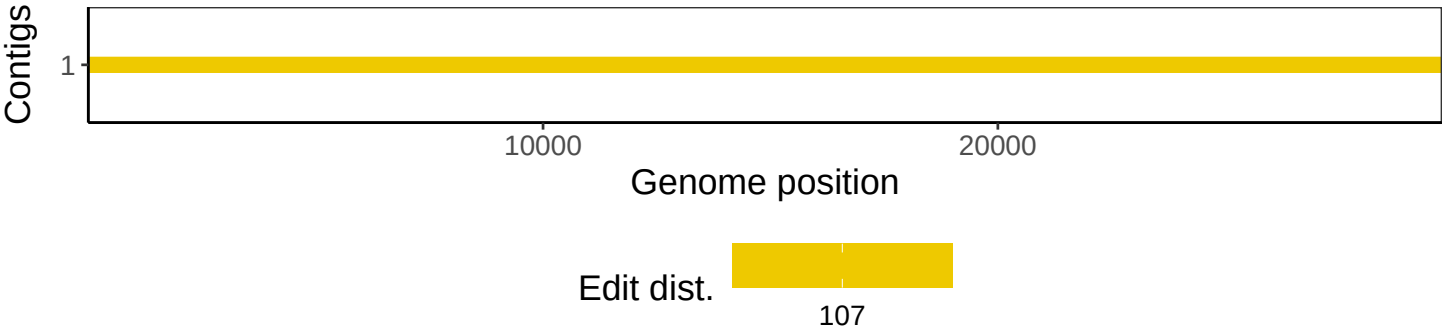
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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