COVID-19 subject UPHS-0438

2021-06-23

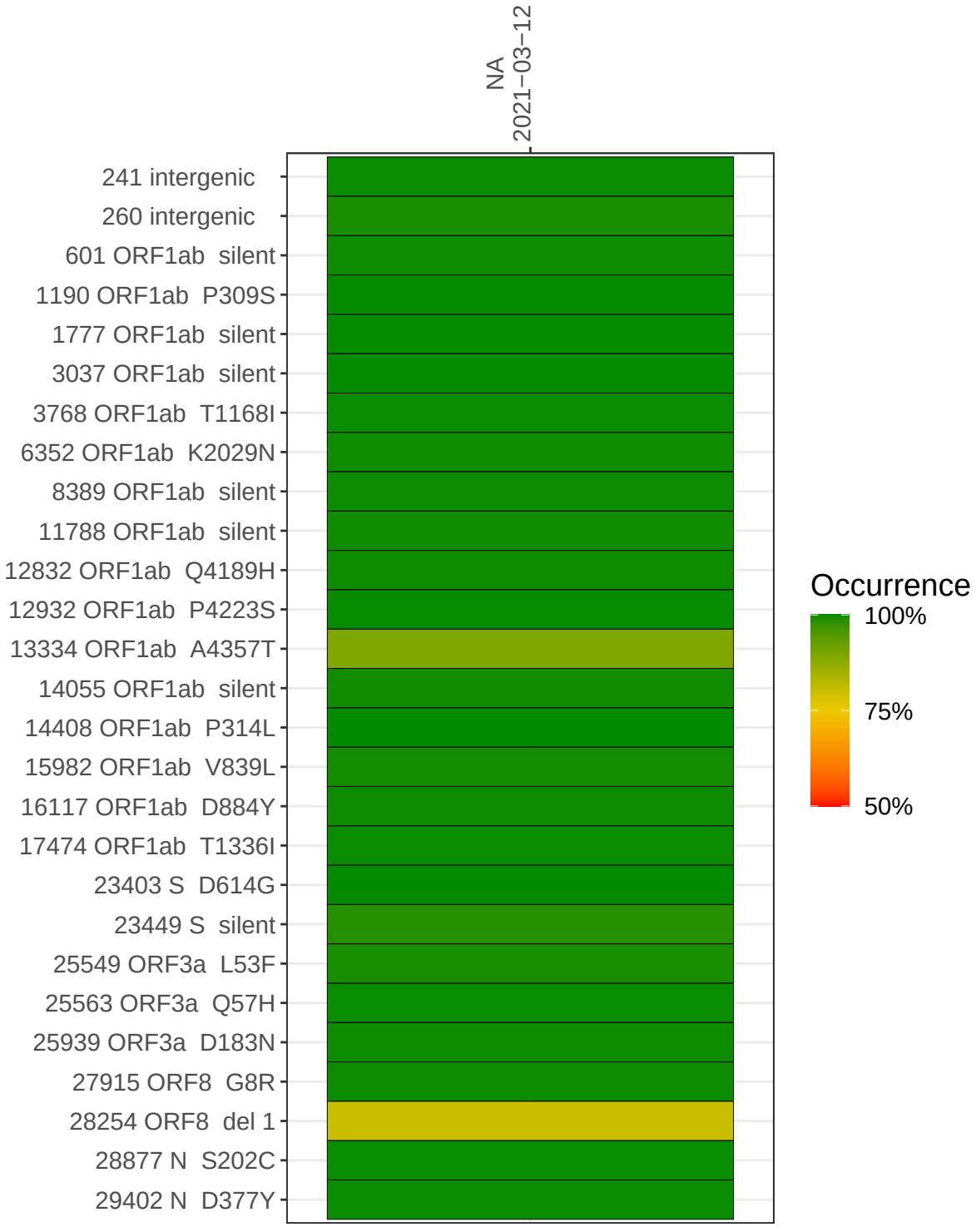
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

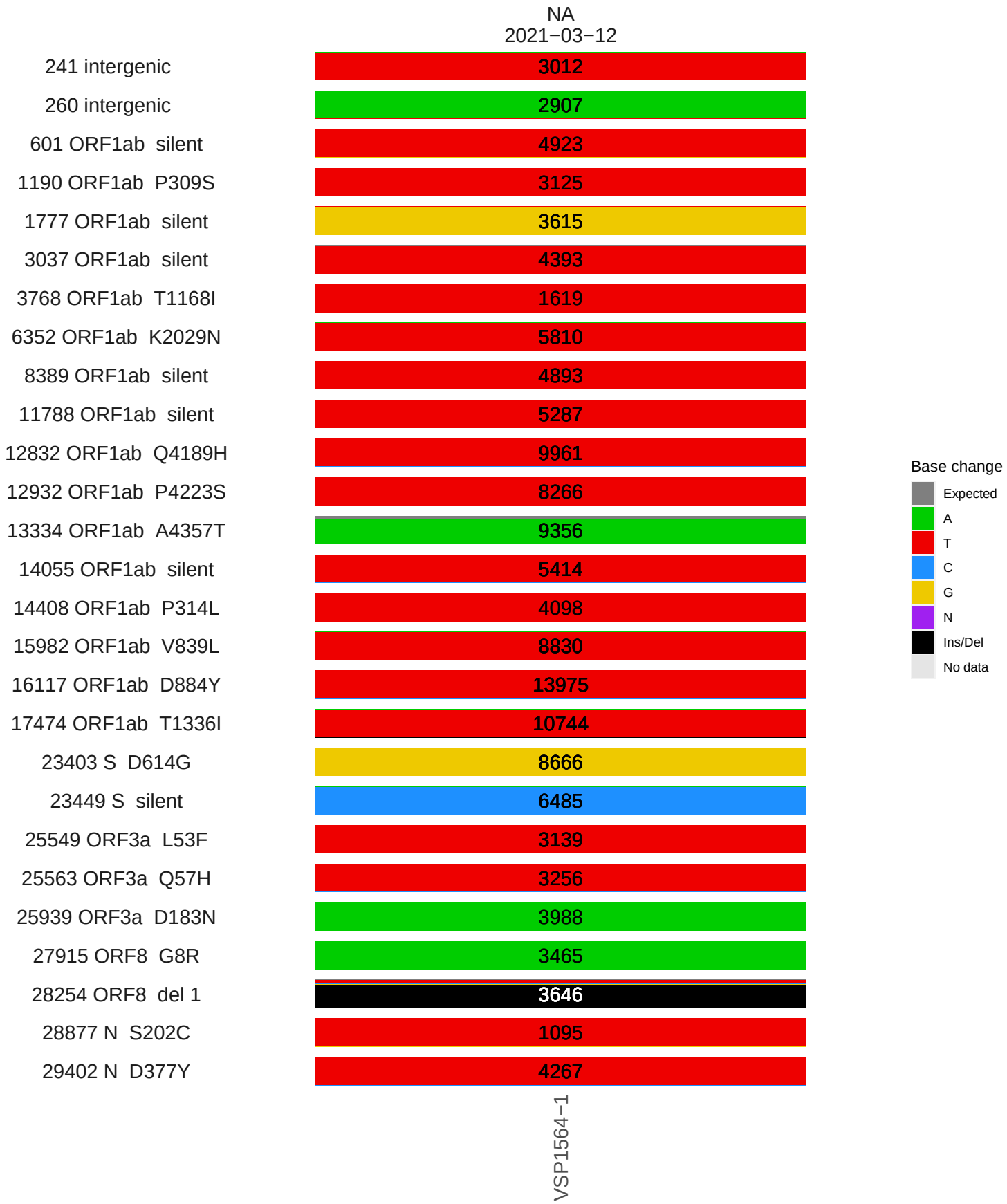
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP1564-1	single experiment	NA	NA	2021-03-12	29.80	B.1.110.3	100.0%	99.8%

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.

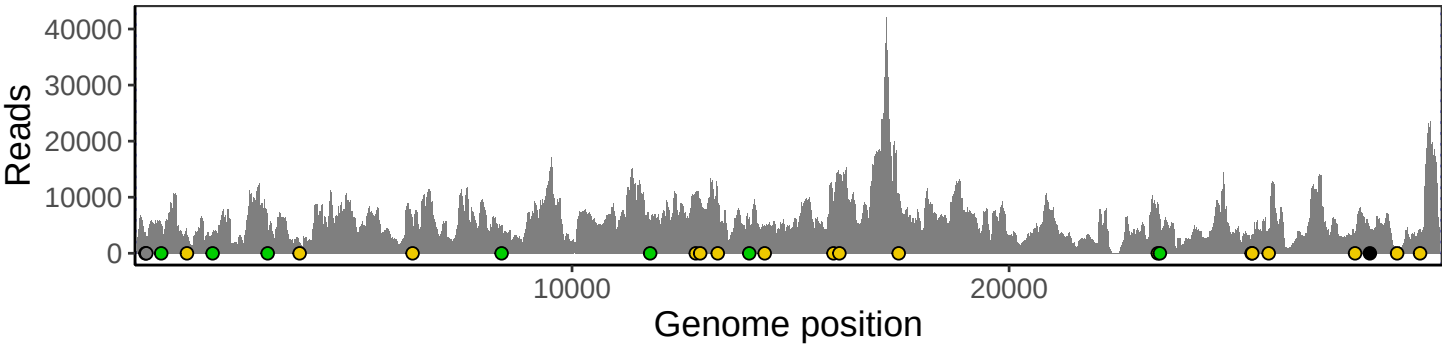



NA
2021−03−12
241 intergenic
3012
260 intergenic
2907
601 ORF1ab silent
4923
1190 ORF1ab P309S
3125
1777 ORF1ab silent
3615
3037 ORF1ab silent
4393
3768 ORF1ab T1168I
1619
6352 ORF1ab K2029N
5810
8389 ORF1ab silent
4893
11788 ORF1ab silent
5287
12832 ORF1ab Q4189H
9961
12932 ORF1ab P4223S
8266
13334 ORF1ab A4357T
9356
14055 ORF1ab silent
5414
14408 ORF1ab P314L
4098
15982 ORF1ab V839L
8830
16117 ORF1ab D884Y
13975
17474 ORF1ab T1336I
10744
23403 S D614G
8666
23449 S silent
6485
25549 ORF3a L53F
3139
25563 ORF3a Q57H
3256
25939 ORF3a D183N
3988
27915 ORF8 G8R
3465
28254 ORF8 del 1
3646
28877 N S202C
1095
29402 N D377Y
4267
VSP1564−1
Base change
Expected
A
T
C
G
N
Ins/Del
No data

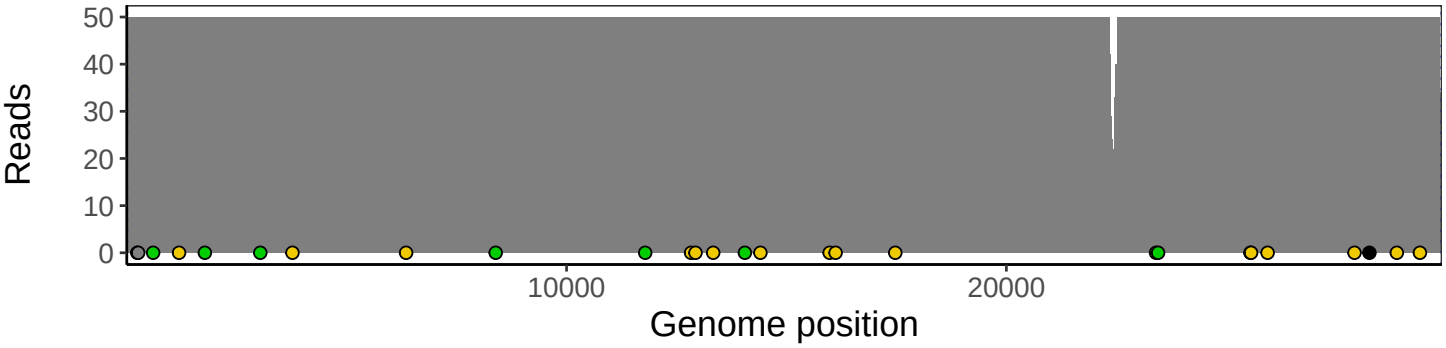
Analyses of individual experiments and composite results

VSP1564-1 | 2021-03-12 | NA | UPHS-0438 | genomes | single experiment

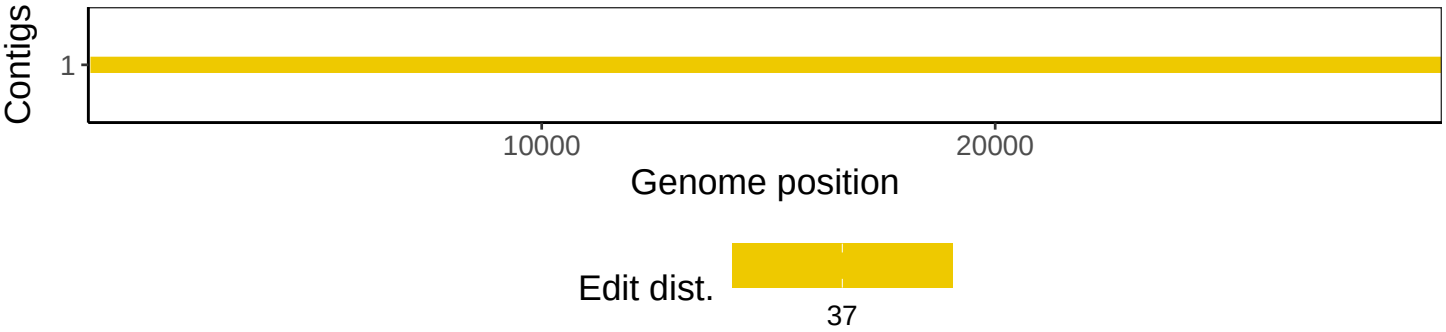
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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