# COVID-19 subject UPHS-0027

2021-03-25

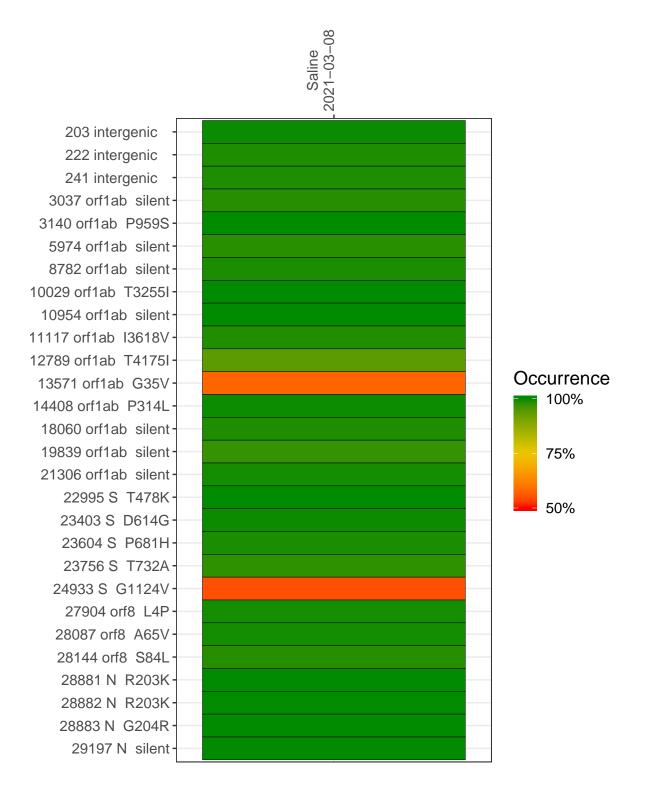
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

Table 1. Sample summary.

| Experiment | Type              | Genomes | Sample type | Sample date | Largest contig<br>(KD) | Lineage   | Reference read<br>coverage | Reference read coverage (>= 5 reads) |
|------------|-------------------|---------|-------------|-------------|------------------------|-----------|----------------------------|--------------------------------------|
| VSP0959-1  | single experiment | NA      | Saline      | 2021-03-08  | 29.83                  | B.1.1.222 | 99.8%                      | 99.8%                                |

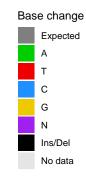
#### Variants shared across samples

The heat map below shows how variants (reference genome 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.



#### Saline 2021-03-08

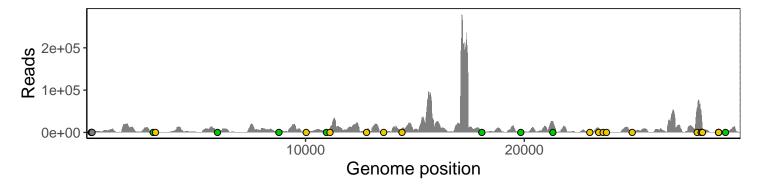
|                     | 2021-03-08   |
|---------------------|--------------|
| 203 intergenic      | 2733         |
| 222 intergenic      | 2660         |
| 241 intergenic      | 2364         |
| 3037 orf1ab silent  | 1200         |
| 3140 orf1ab P959S   | 571          |
| 5974 orf1ab silent  | 1135         |
| 8782 orf1ab silent  | 5321         |
| 10029 orf1ab T3255I | 554          |
| 10954 orf1ab silent | 3282         |
| 11117 orf1ab I3618V | 9898         |
| 12789 orf1ab T4175I | <b>70</b> 65 |
| 13571 orf1ab G35V   | 947          |
| 14408 orf1ab P314L  | 9705         |
| 18060 orf1ab silent | 693          |
| 19839 orf1ab silent | 3064         |
| 21306 orf1ab silent | 23986        |
| 22995 S T478K       | 189          |
| 23403 S D614G       | 10855        |
| 23604 S P681H       | 4533         |
| 23756 S T732A       | 2726         |
| 24933 S G1124V      | 5505         |
| 27904 orf8 L4P      | 56696        |
| 28087 orf8 A65V     | 46086        |
| 28144 orf8 S84L     | 10341        |
| 28881 N R203K       | 94           |
| 28882 N R203K       | 94           |
| 28883 N G204R       | 94           |
| 29197 N silent      | 3619         |
|                     | 1959–1       |
|                     | 902          |



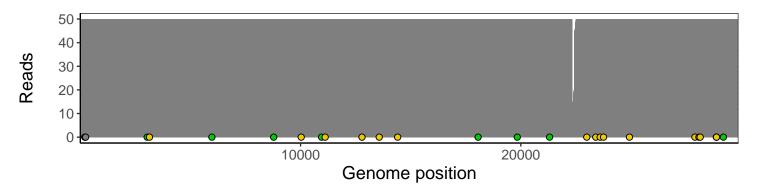
### Analyses of individual experiments and composite results

#### $VSP0959\text{-}1 \mid 2021\text{-}03\text{-}08 \mid Saline \mid UPHS\text{-}0027 \mid genomes \mid single \ experiment$

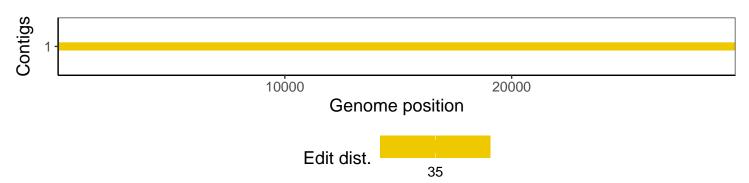
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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                           | 2.3.3  |
| 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                            | 3.0.0  |
| tidyverse                          | 1.2.1  |
| ShortRead                          | 1.34.2   |
| $\operatorname{GenomicAlignments}$ | 1.12.2   |
| ${\bf Summarized Experiment}$      | 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   |