

COVID-19 subject UPHS-0023

2021-04-17

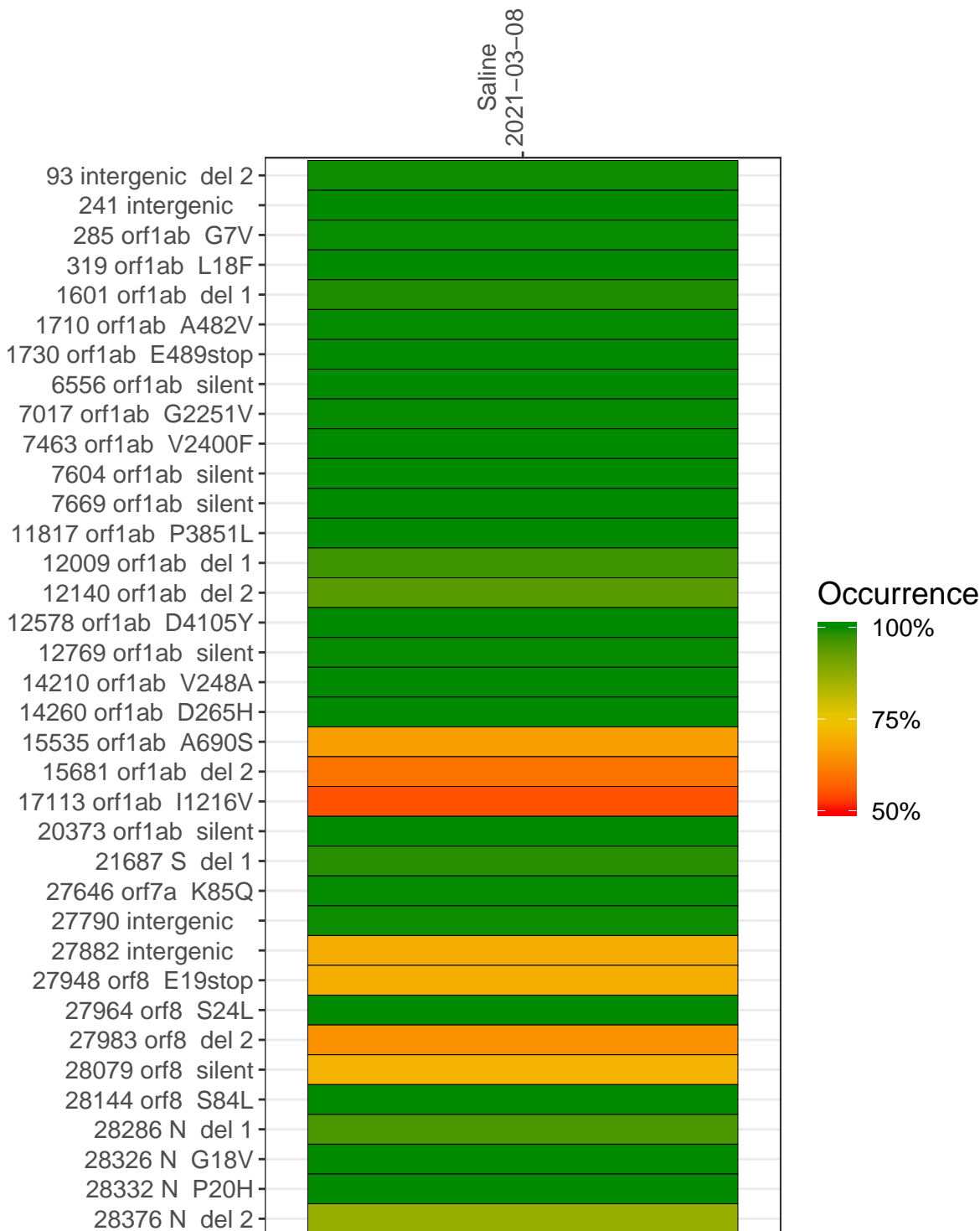
The table below provides a summary of subject samples for which sequencing data is available. The experiments column shows the number of sequencing experiments performed for each specimen. Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software 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)
VSP0955-1	single experiment	NA	Saline	2021-03-08	1.03	NA	26.0%	23.7%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/SARS-CoV-2-Philadelphia/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	
93 intergenic del 2	602	
241 intergenic	783	
285 orf1ab G7V	634	
319 orf1ab L18F	844	
1601 orf1ab del 1	972	
1710 orf1ab A482V	5860	
1730 orf1ab E489stop	6624	
6556 orf1ab silent	4141	
7017 orf1ab G2251V	2656	
7463 orf1ab V2400F	8340	
7604 orf1ab silent	7128	
7669 orf1ab silent	1219	
11817 orf1ab P3851L	9040	
12009 orf1ab del 1	8150	
12140 orf1ab del 2	5626	
12578 orf1ab D4105Y	10187	
12769 orf1ab silent	5204	
14210 orf1ab V248A	5779	
14260 orf1ab D265H	2493	
15535 orf1ab A690S	5862	
15681 orf1ab del 2	9462	
17113 orf1ab I1216V	10999	
20373 orf1ab silent	3795	
21687 S del 1	2643	
27646 orf7a K85Q	2962	
27790 intergenic	1883	
27882 intergenic	9226	
27948 orf8 E19stop	10819	
27964 orf8 S24L	10928	
27983 orf8 del 2	9529	
28079 orf8 silent	7864	
28144 orf8 S84L	4233	
28286 N del 1	3076	
28326 N G18V	2820	
28332 N P20H	2735	
28376 N del 2	2209	

Base change

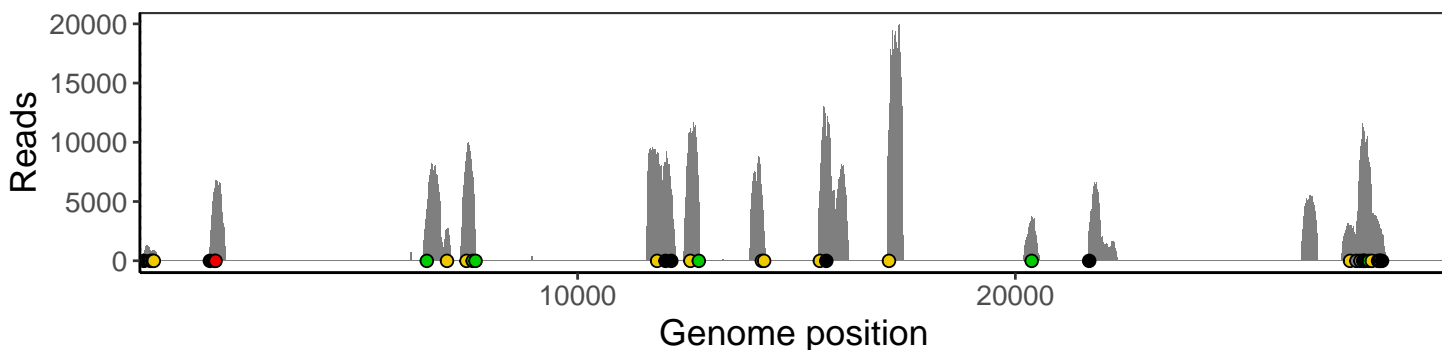
- Expected
- A
- T
- C
- G
- N
- Ins/Del
- No data

VSP0955-1

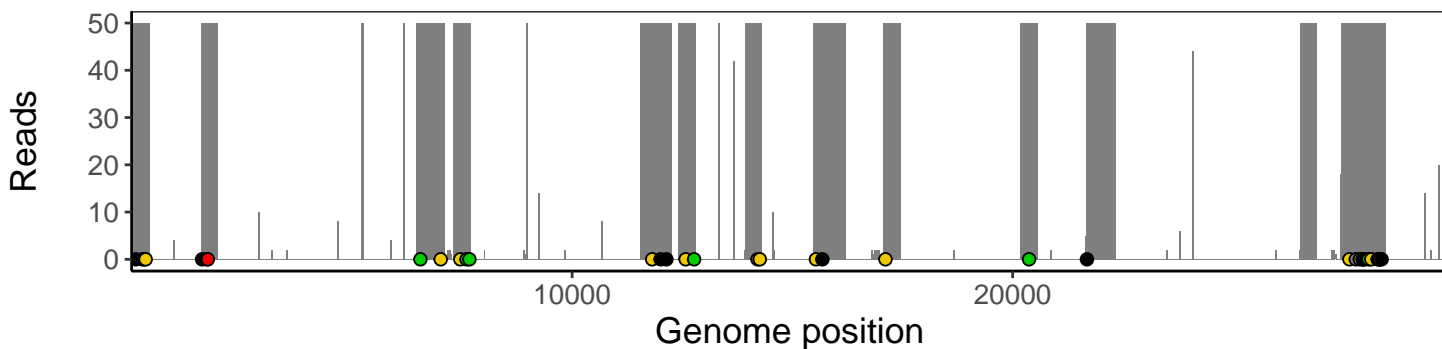
Analyses of individual experiments and composite results

VSP0955-1 | 2021-03-08 | Saline | UPHS-0023 | genomes | single experiment

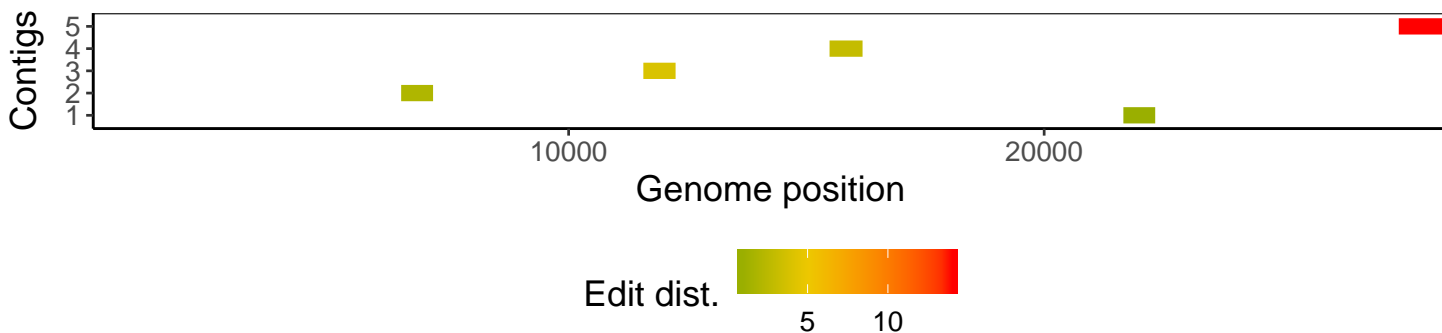
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according to variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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 htlib 1.10
bcftools	1.10.2-34-g1a12af0-dirty Using htlib 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
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