COVID-19 subject HUP-PH-0004

2021-05-05

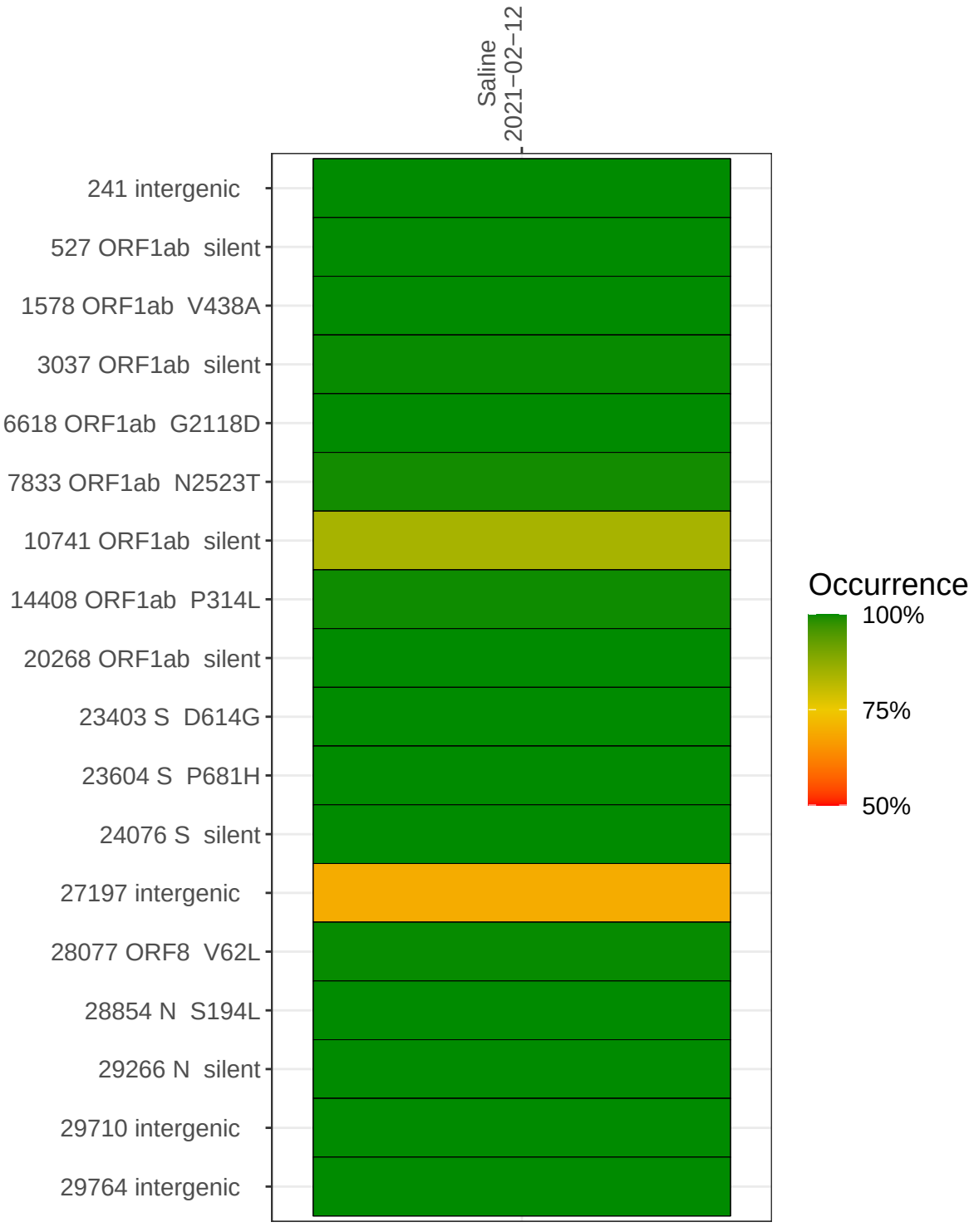
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

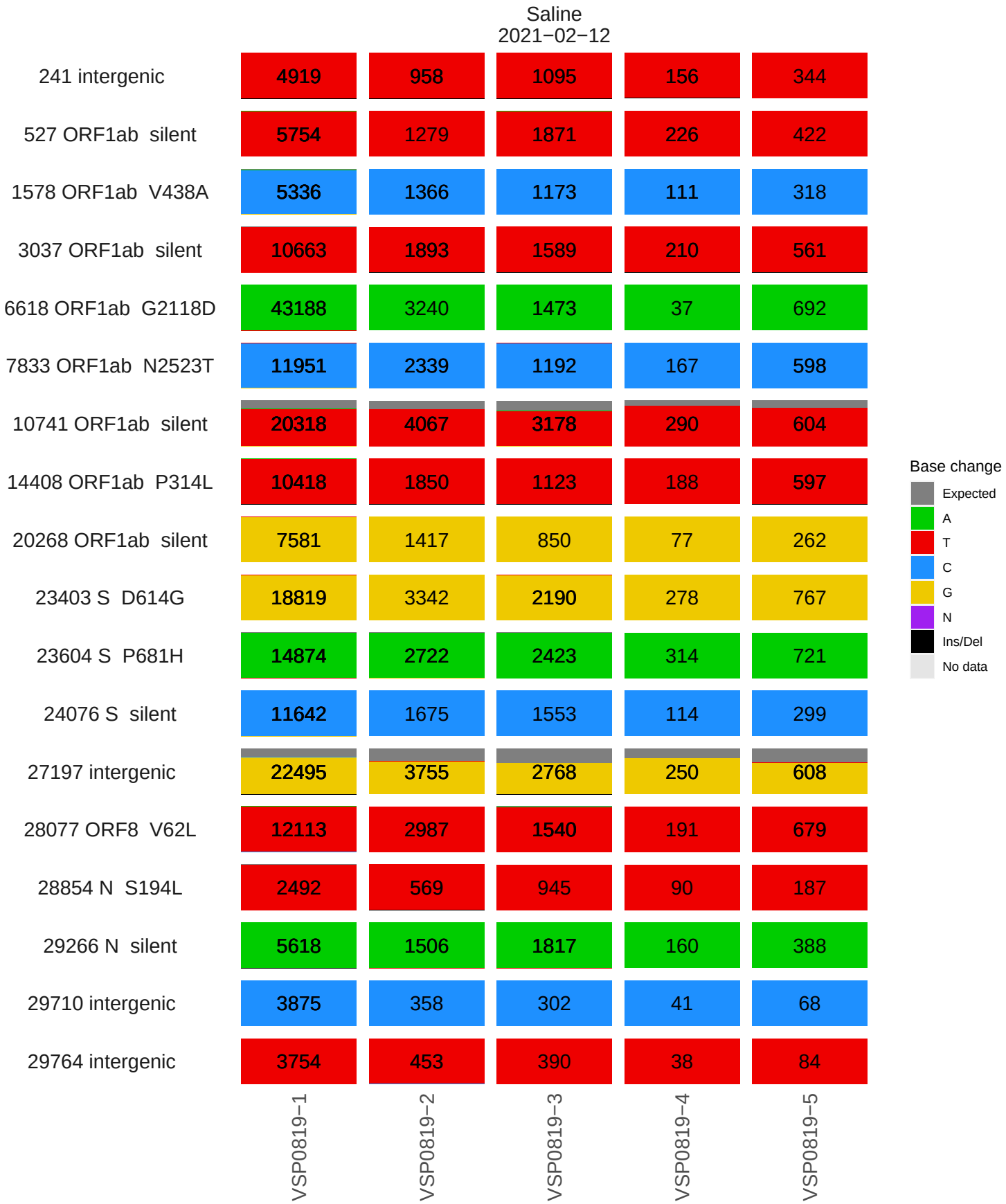
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0819-1	single experiment	NA	Saline	2021-02-12	29.89	B.1.243	100.0%	99.9%
VSP0819-2	single experiment	NA	Saline	2021-02-12	29.86	B.1.243	99.8%	99.7%
VSP0819-3	single experiment	NA	Saline	2021-02-12	29.87	B.1.243	99.8%	99.7%
VSP0819-4	single experiment	NA	Saline	2021-02-12	29.80	B.1.243	99.7%	99.6%
VSP0819-5	single experiment	NA	Saline	2021-02-12	29.82	B.1.243	99.7%	99.7%

Variants shared across samples

The heat map below shows how variants (reference genome /home/everett/projects/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.

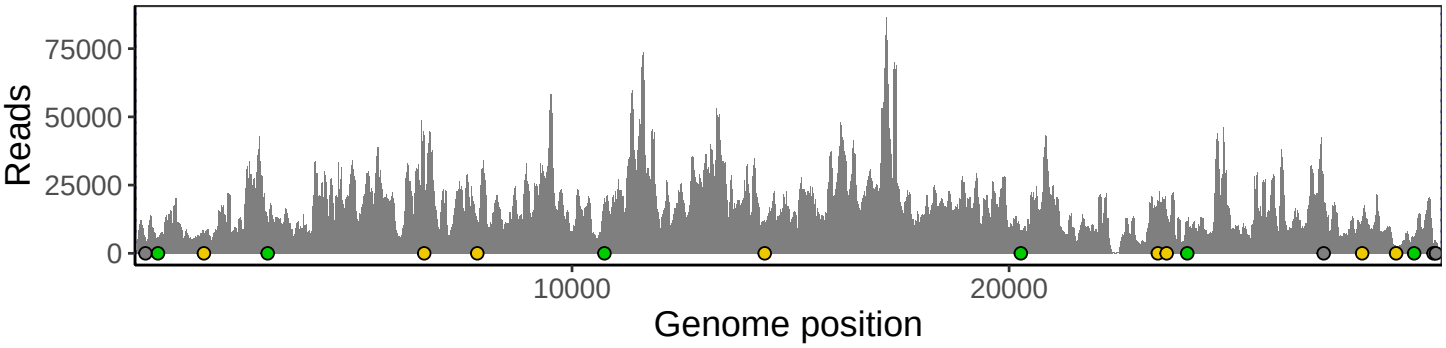



Saline
2021−02−12
241 intergenic
527 ORF1ab silent
1578 ORF1ab V438A
3037 ORF1ab silent
6618 ORF1ab G2118D
7833 ORF1ab N2523T
10741 ORF1ab silent
14408 ORF1ab P314L
20268 ORF1ab silent
23403 S D614G
23604 S P681H
24076 S silent
27197 intergenic
28077 ORF8 V62L
28854 N S194L
29266 N silent
29710 intergenic
29764 intergenic
4919 958 1095 156 344
5754 1279 1871 226 422
5336 1366 1173 111 318
10663 1893 1589 210 561
43188 3240 1473 37 692
11951 2339 1192 167 598
20318 4067 3178 290 604
10418 1850 1123 188 597
7581 1417 850 77 262
18819 3342 2190 278 767
14874 2722 2423 314 721
11642 1675 1553 114 299
22495 3755 2768 250 608
12113 2987 1540 191 679
2492 569 945 90 187
5618 1506 1817 160 388
3875 358 302 41 68
3754 453 390 38 84
VSP0819−1
VSP0819−2
VSP0819−3
VSP0819−4
VSP0819−5
Base change
Expected
A
T
C
G
N
Ins/Del
No data

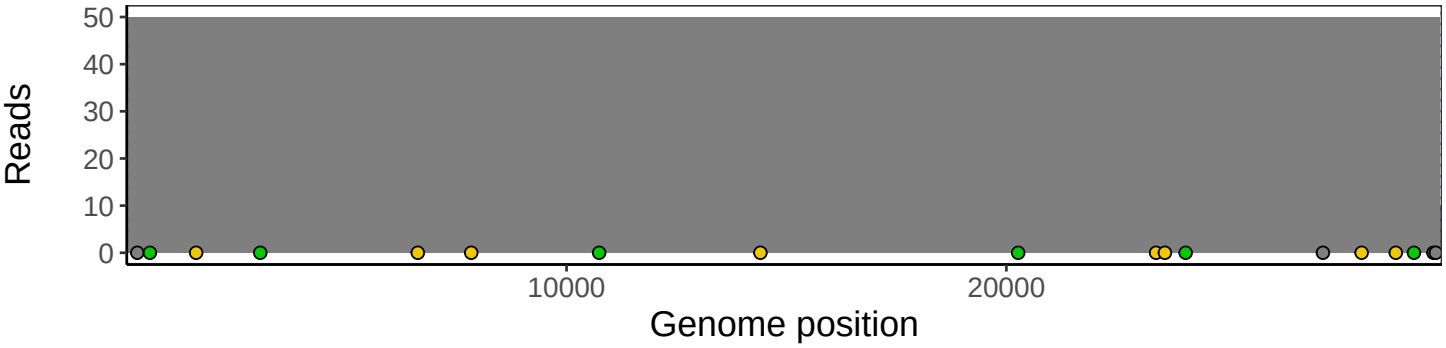
Analyses of individual experiments and composite results

VSP0819-1 | 2021-02-12 | Saline | HUP-PH-0004 | genomes | single experiment

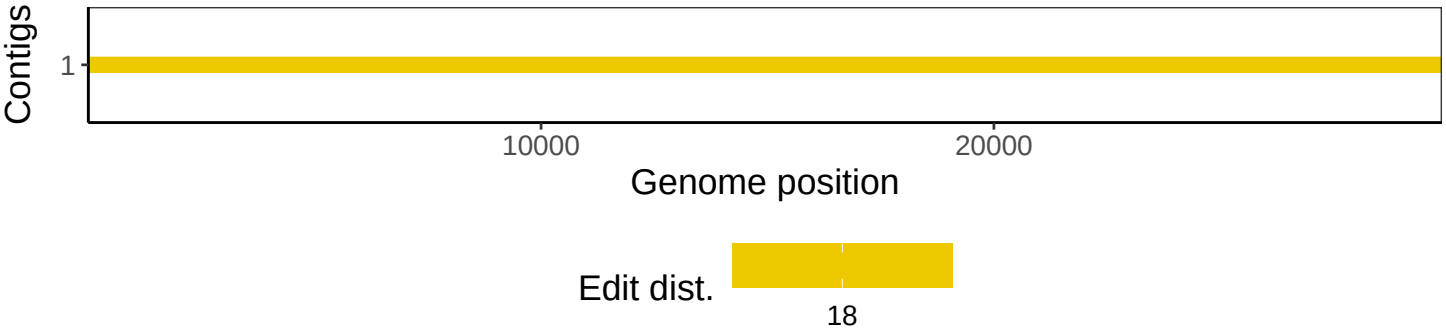
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

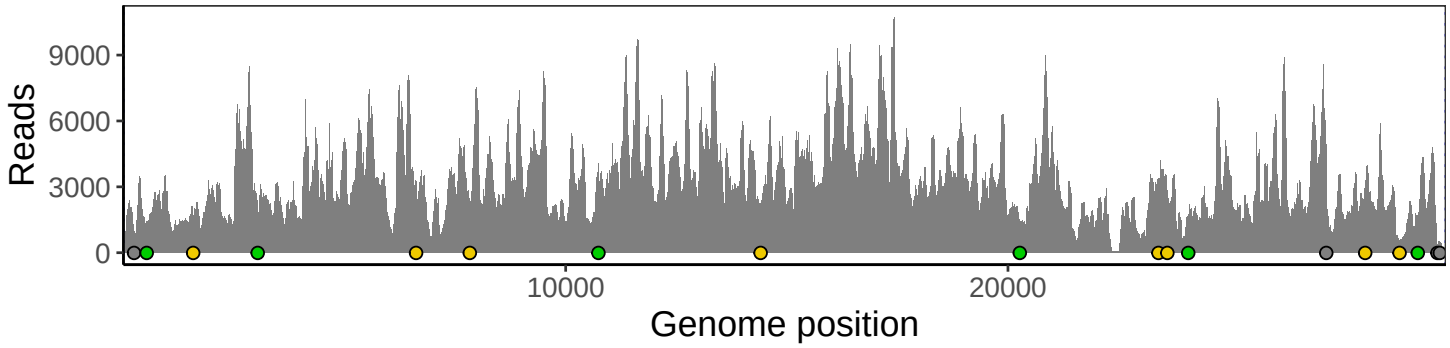


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

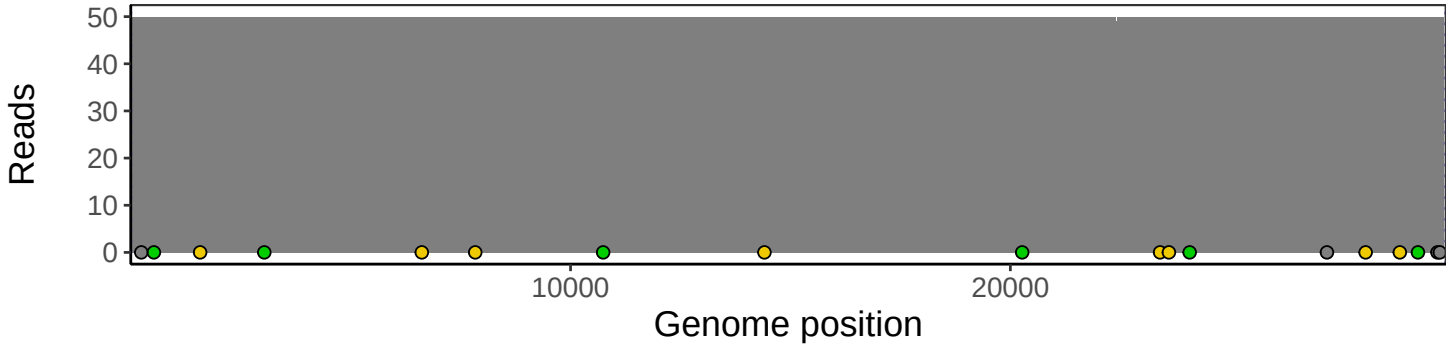


VSP0819-2 | 2021-02-12 | Saline | HUP-PH-0004 | genomes | single experiment

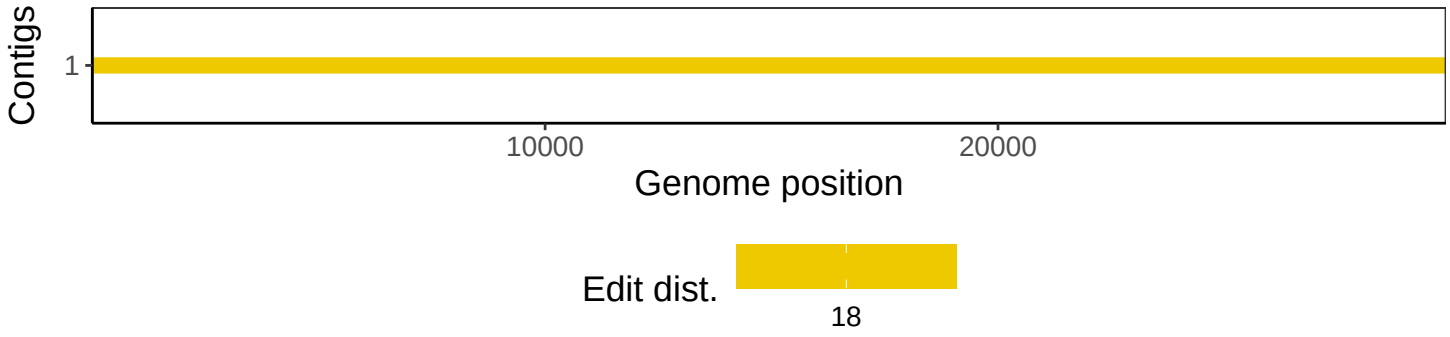
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

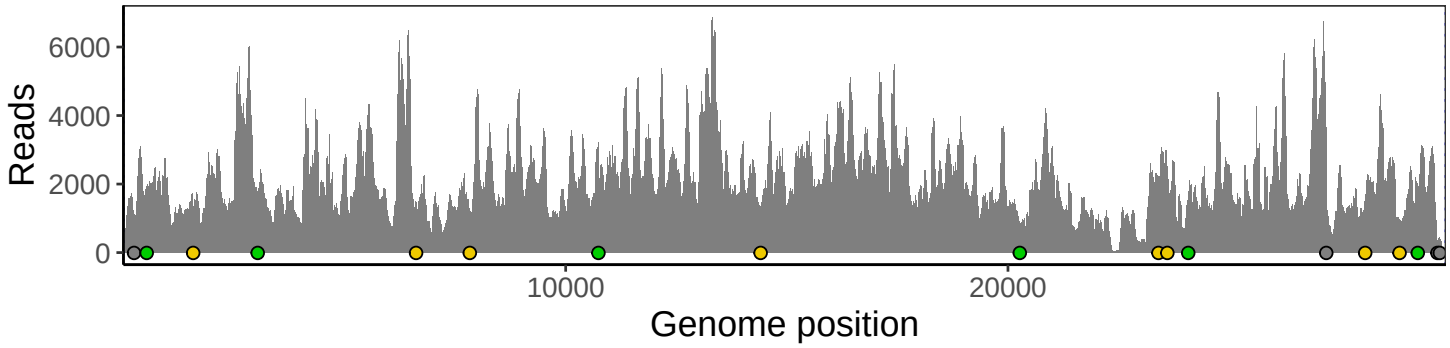


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

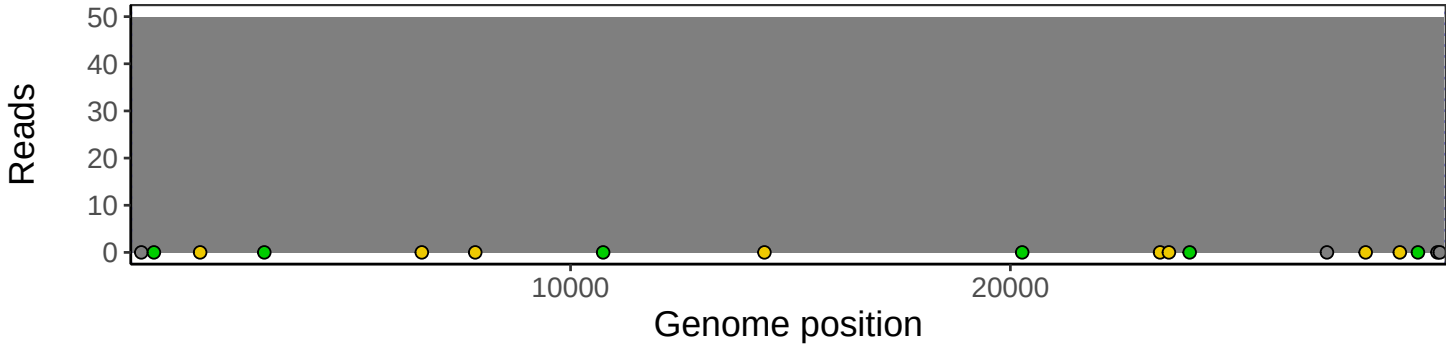


VSP0819-3 | 2021-02-12 | Saline | HUP-PH-0004 | genomes | single experiment

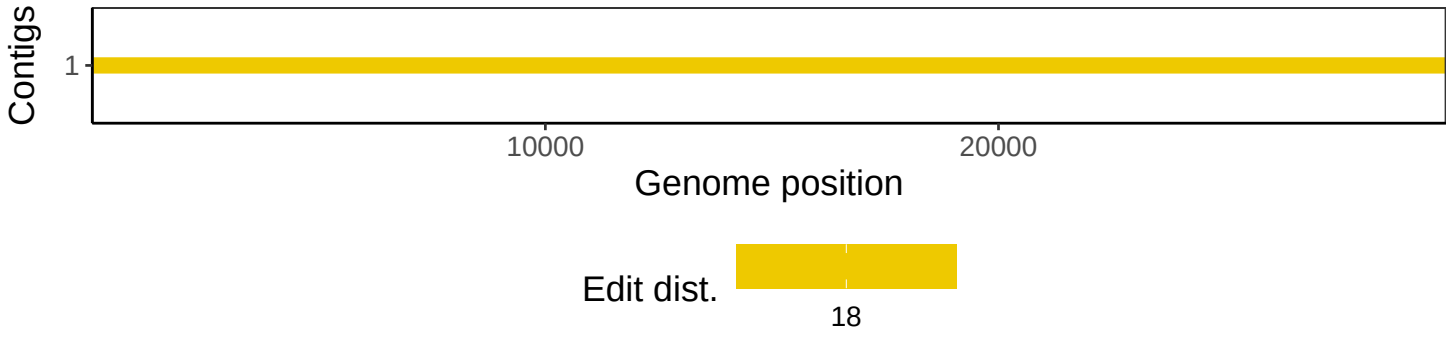
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

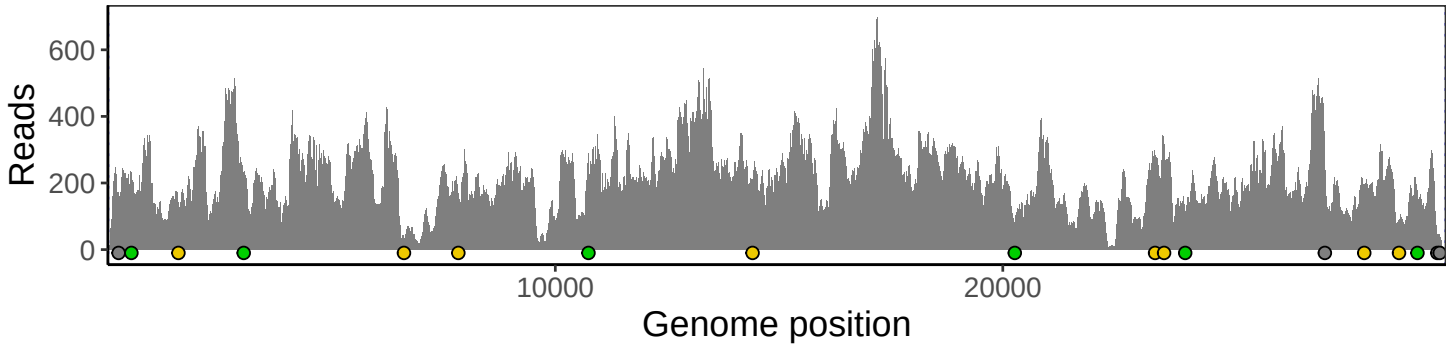


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

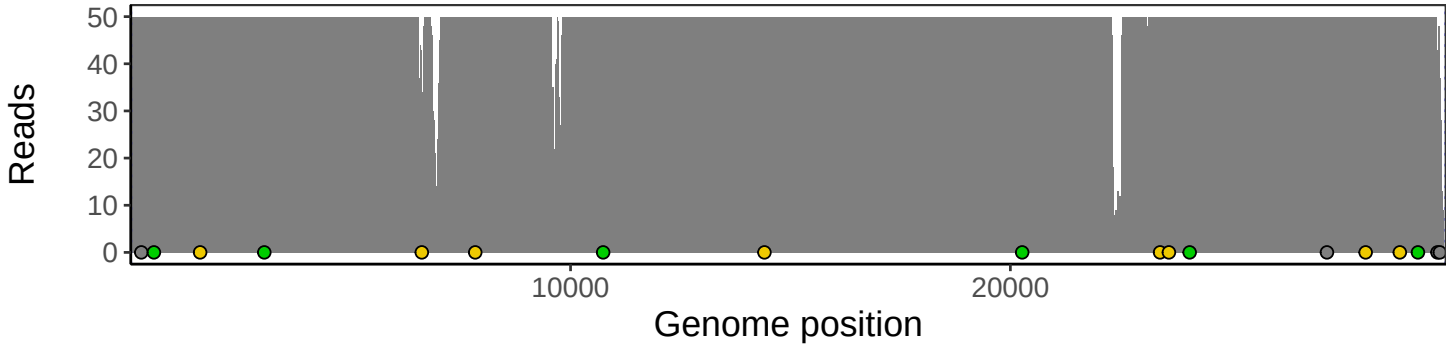


VSP0819-4 | 2021-02-12 | Saline | HUP-PH-0004 | genomes | single experiment

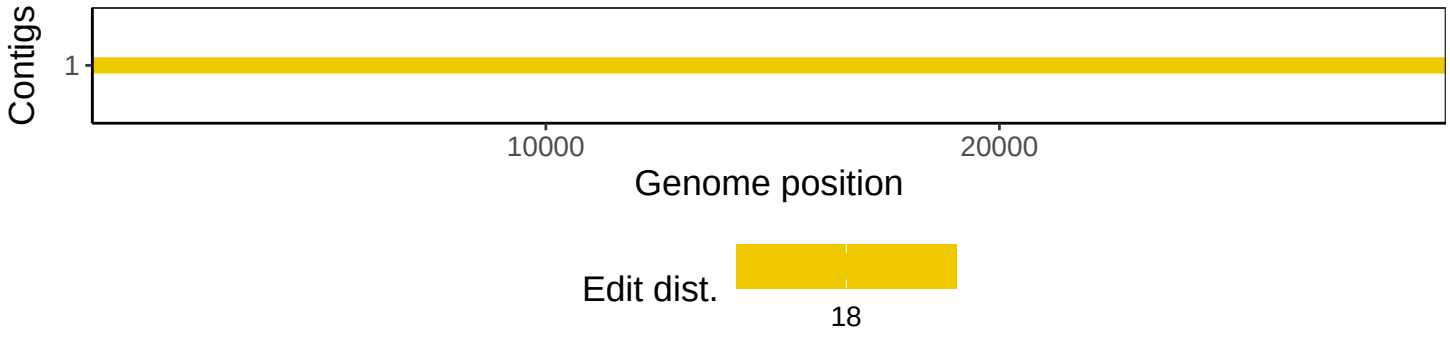
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



Excerpt from plot above focusing on reads coverage from 0 to 50 NT.

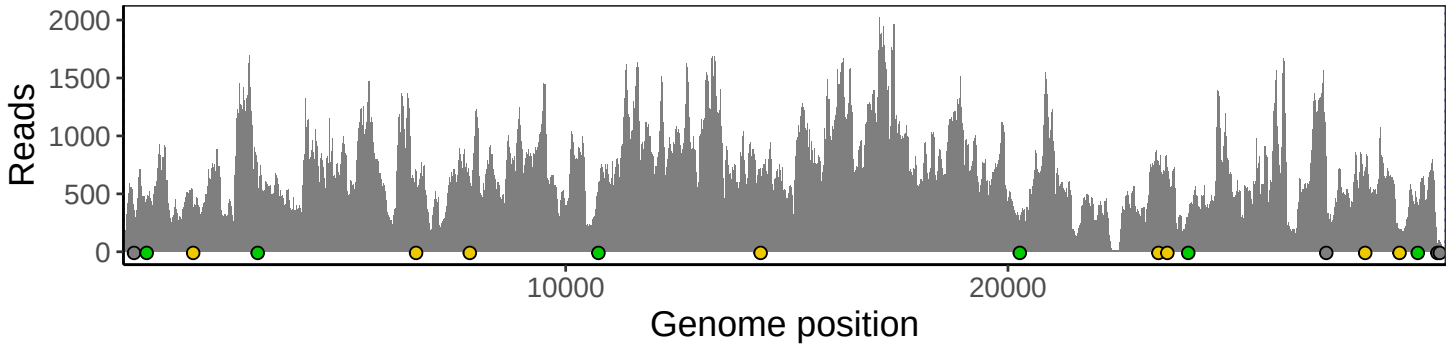


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

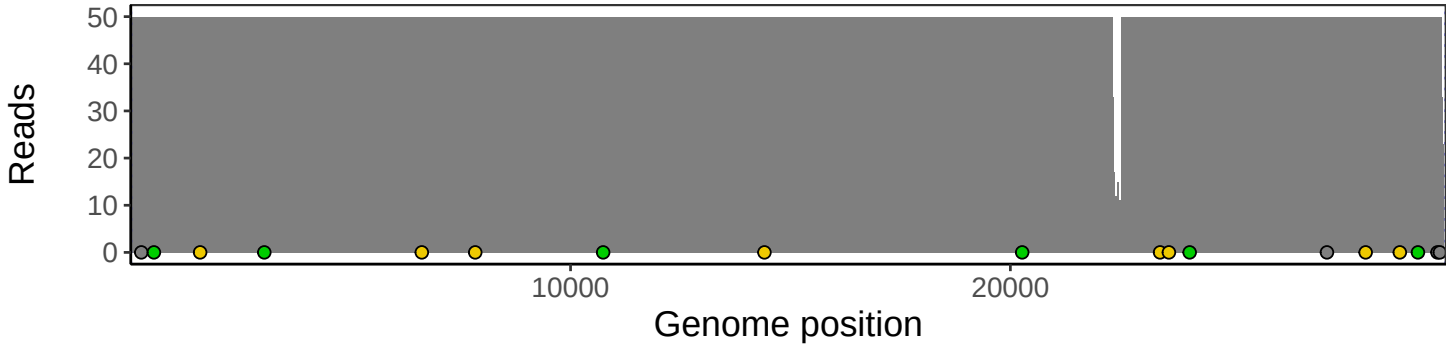


VSP0819-5 | 2021-02-12 | Saline | HUP-PH-0004 | genomes | single experiment

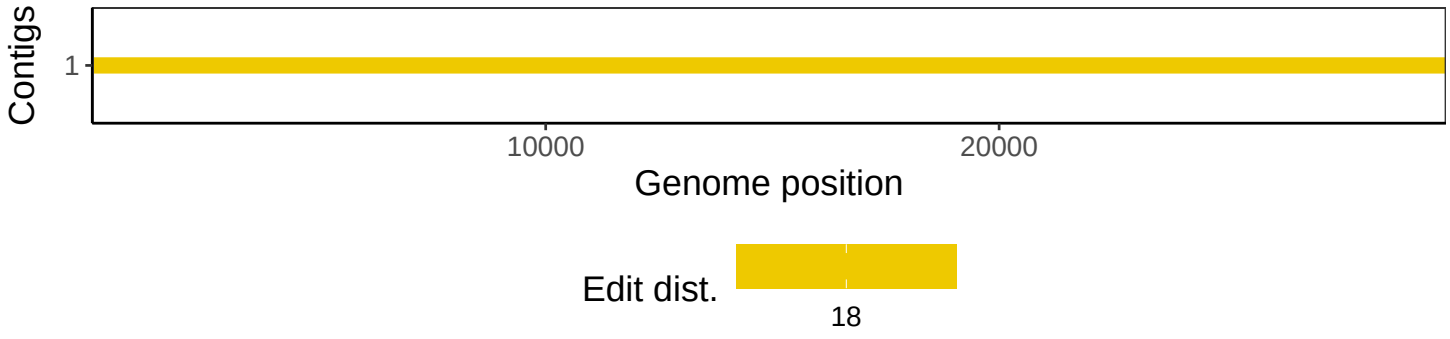
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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	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