COVID-19 subject MPCluster2-Seq14

2021-05-05

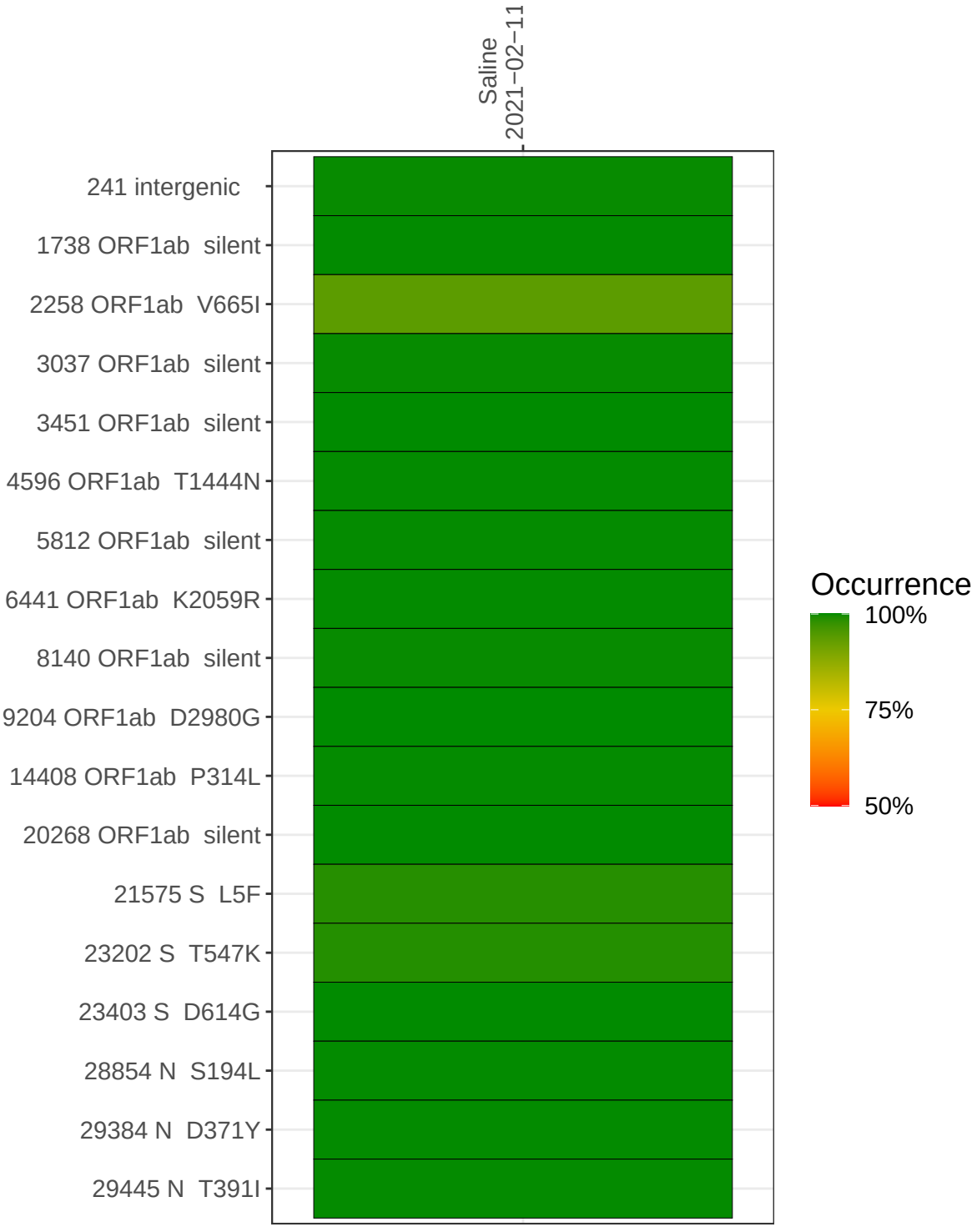
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report. Lineages are called with the Pangolin software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

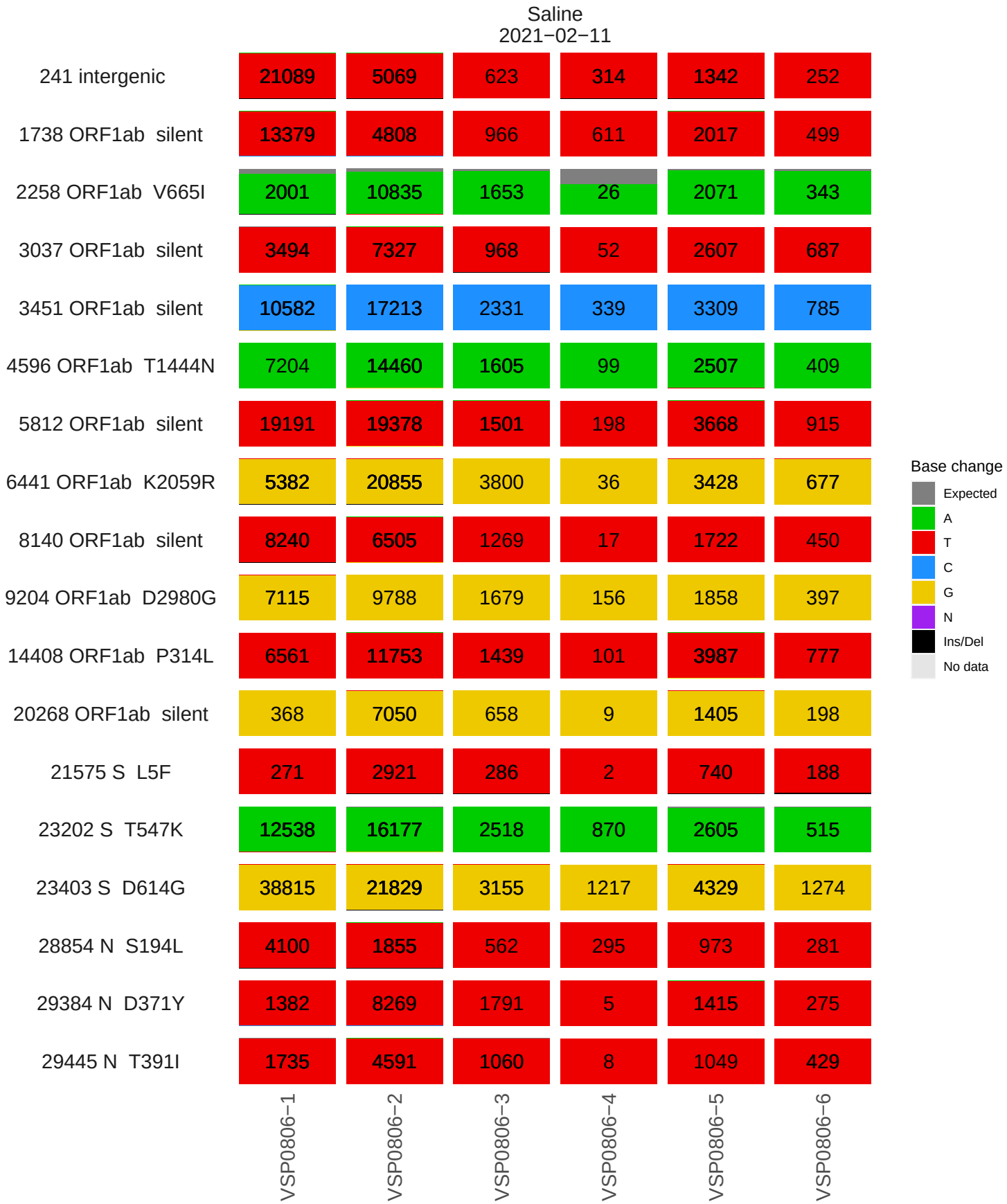
Table 1. Sample summary.

Experiment	Type	Genomes	Sample type	Sample date	Largest contig (KD)	Lineage	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0806	composite	NA	Saline	2021-02-11	25.39	B.1.234	100.0%	99.9%
VSP0806-1	single experiment	NA	Saline	2021-02-11	29.47	B.1.234	100.0%	99.8%
VSP0806-2	single experiment	NA	Saline	2021-02-11	29.85	B.1.234	100.0%	99.8%
VSP0806-3	single experiment	NA	Saline	2021-02-11	29.84	B.1.234	99.7%	99.7%
VSP0806-4	single experiment	NA	Saline	2021-02-11	11.75	B.1.234	99.3%	96.0%
VSP0806-5	single experiment	NA	Saline	2021-02-11	29.91	B.1.234	99.7%	99.7%
VSP0806-6	single experiment	NA	Saline	2021-02-11	29.82	B.1.234	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.

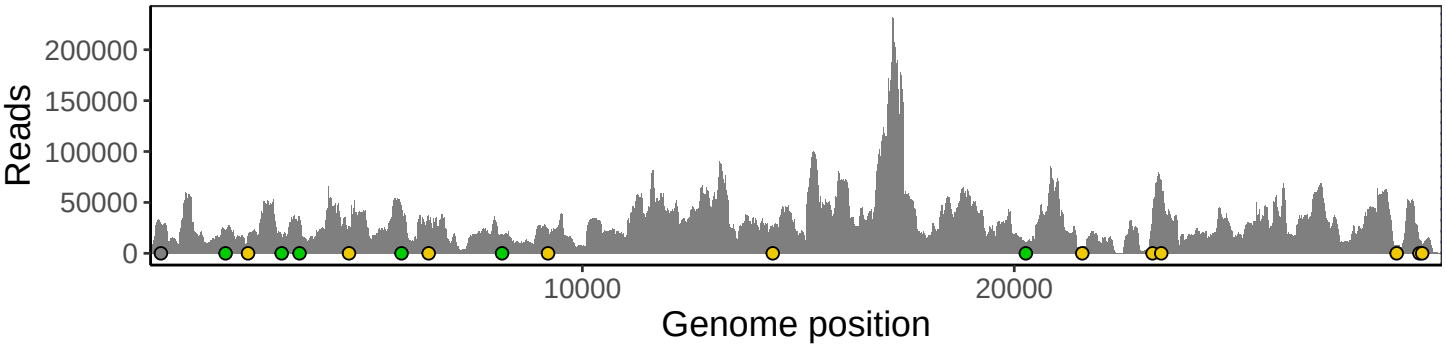



Saline
2021−02−11
241 intergenic
1738 ORF1ab silent
2258 ORF1ab V665I
3037 ORF1ab silent
3451 ORF1ab silent
4596 ORF1ab T1444N
5812 ORF1ab silent
6441 ORF1ab K2059R
8140 ORF1ab silent
9204 ORF1ab D2980G
14408 ORF1ab P314L
20268 ORF1ab silent
21575 S L5F
23202 S T547K
23403 S D614G
28854 N S194L
29384 N D371Y
29445 N T391I
21089 5069 623 314 1342 252
13379 4808 966 611 2017 499
2001 10835 1653 26 2071 343
3494 7327 968 52 2607 687
10582 17213 2331 339 3309 785
7204 14460 1605 99 2507 409
19191 19378 1501 198 3668 915
5382 20855 3800 36 3428 677
8240 6505 1269 17 1722 450
7115 9788 1679 156 1858 397
6561 11753 1439 101 3987 777
368 7050 658 9 1405 198
271 2921 286 2 740 188
12538 16177 2518 870 2605 515
38815 21829 3155 1217 4329 1274
4100 1855 562 295 973 281
1382 8269 1791 5 1415 275
1735 4591 1060 8 1049 429
VSP0806−1
VSP0806−2
VSP0806−3
VSP0806−4
VSP0806−5
VSP0806−6
Base change
Expected
A
T
C
G
N
Ins/Del
No data

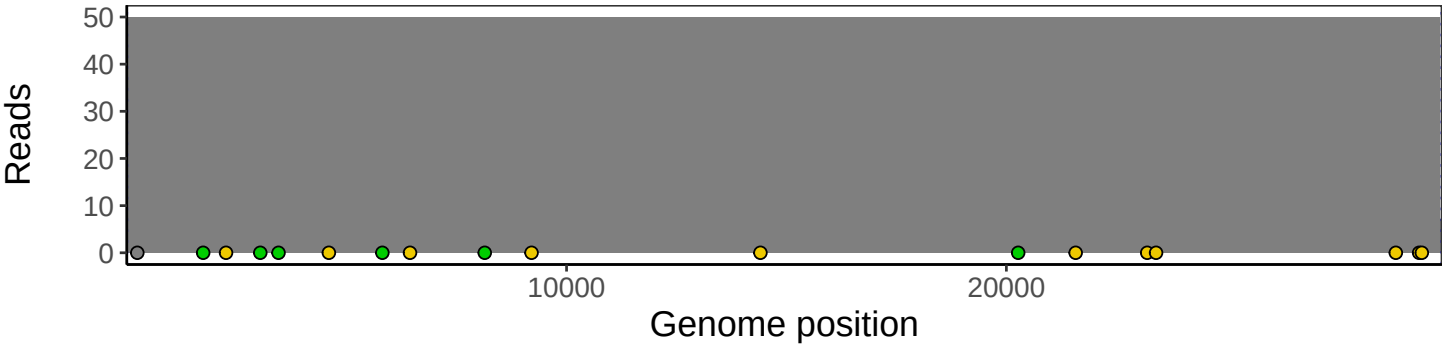
Analyses of individual experiments and composite results

VSP0806 | 2021-02-11 | Saline | MPCluster2-Seq14 | composite result

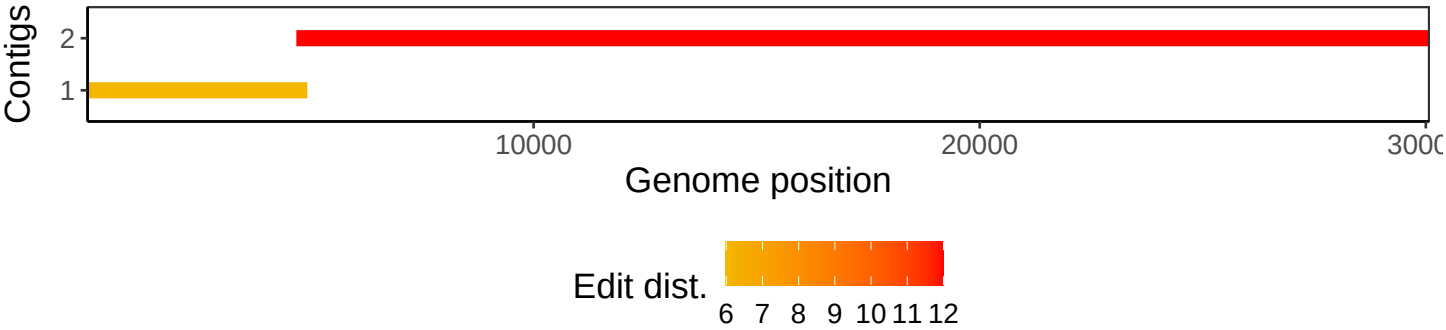
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

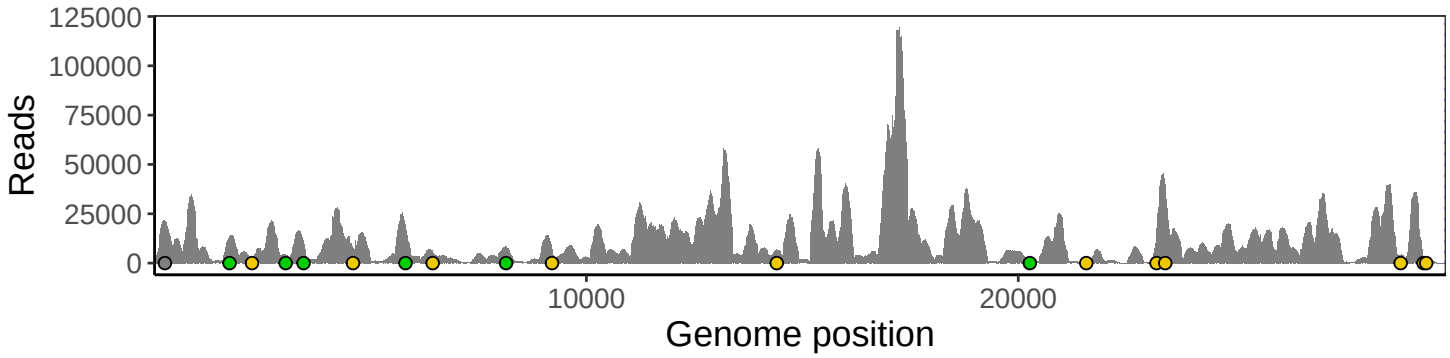


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

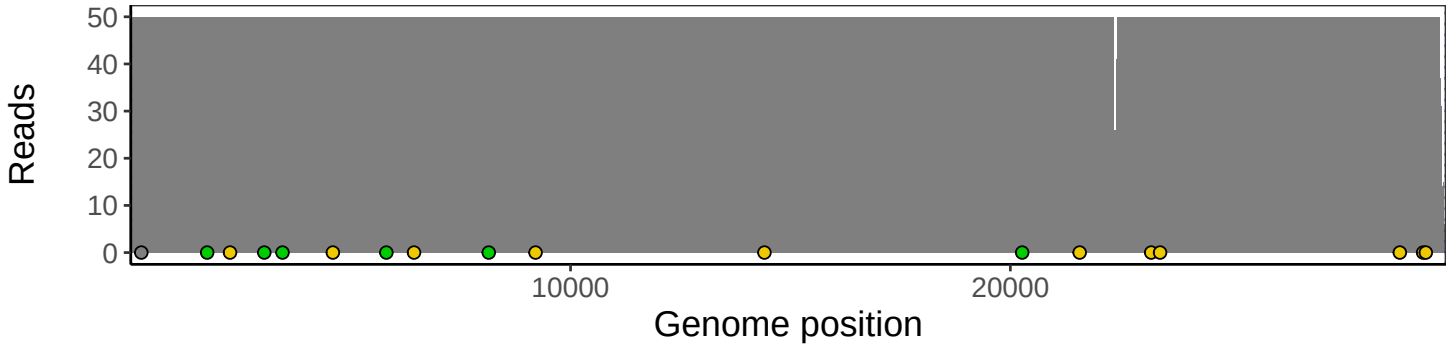


VSP0806-1 | 2021-02-11 | Saline | MPCluster2-Seq14 | genomes | single experiment

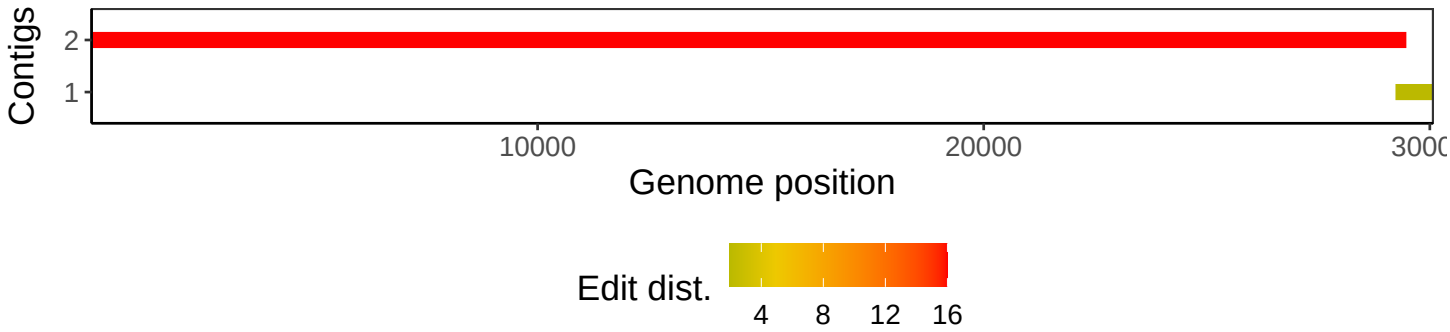
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

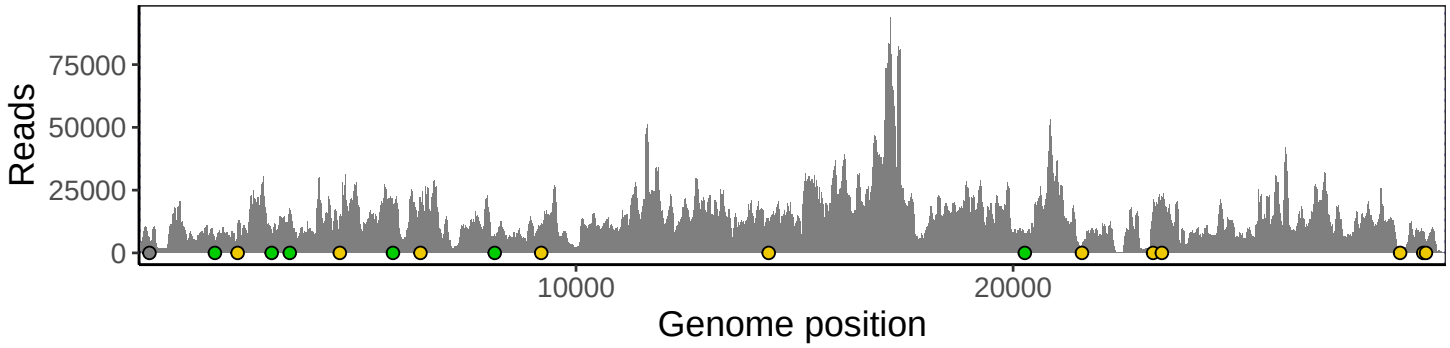


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

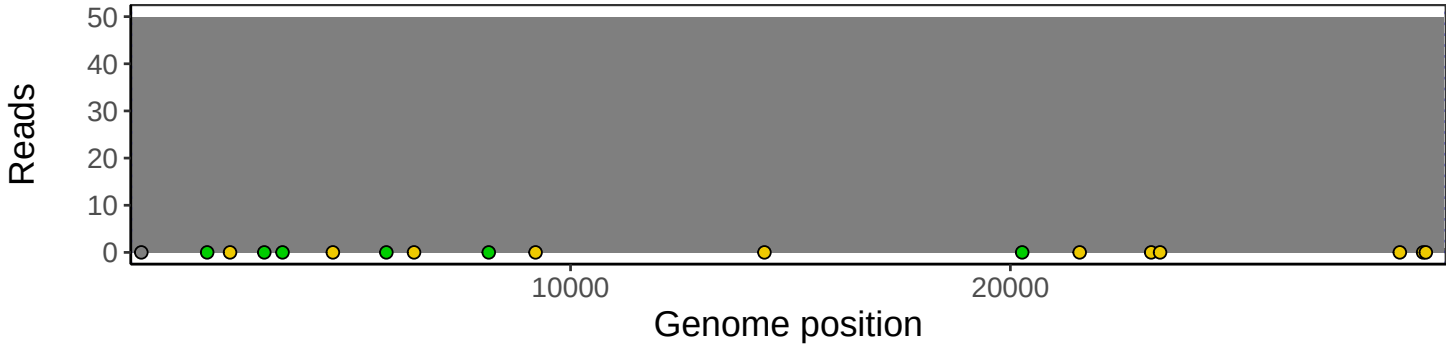


VSP0806-2 | 2021-02-11 | Saline | MPCluster2-Seq14 | genomes | single experiment

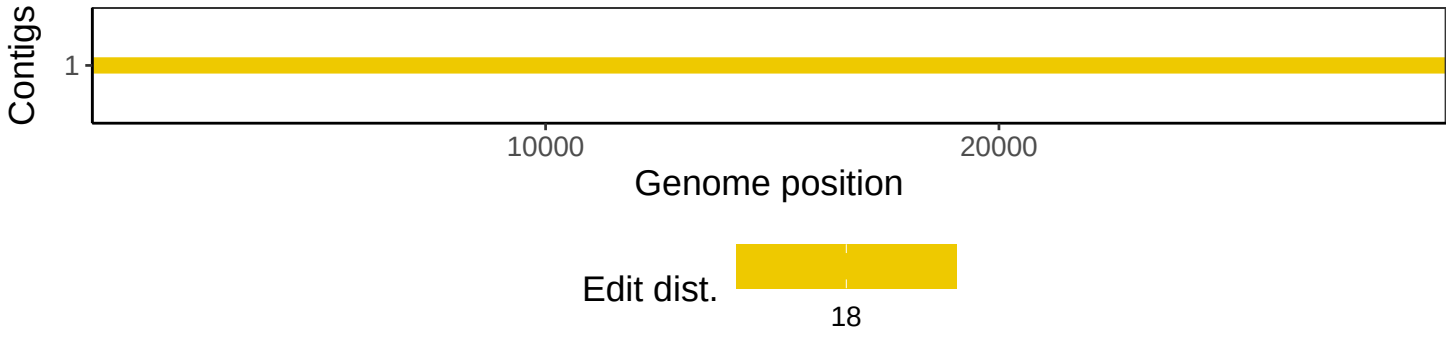
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

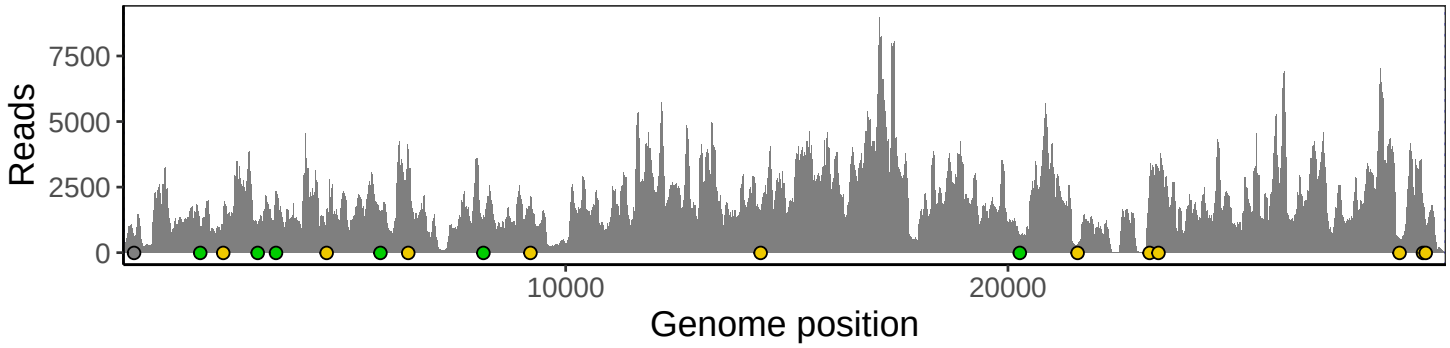


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

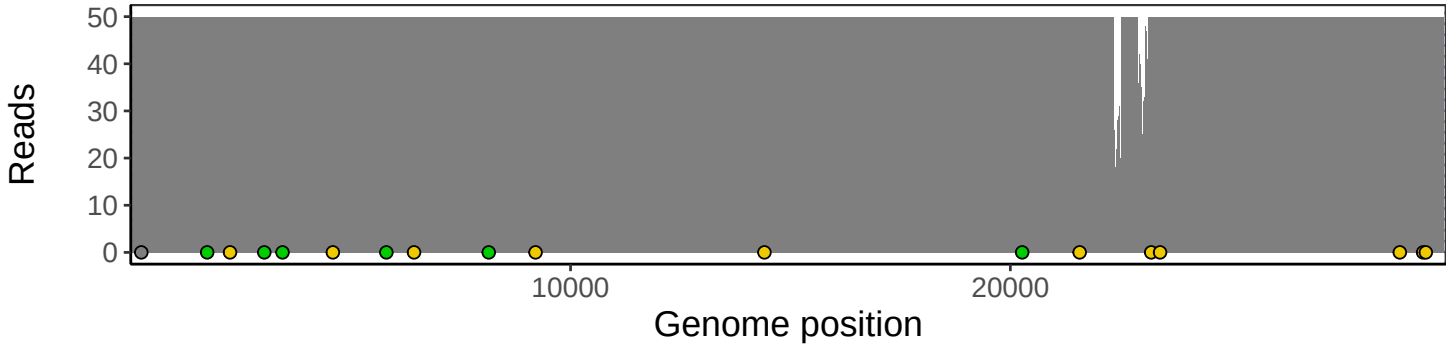


VSP0806-3 | 2021-02-11 | Saline | MPCluster2-Seq14 | genomes | single experiment

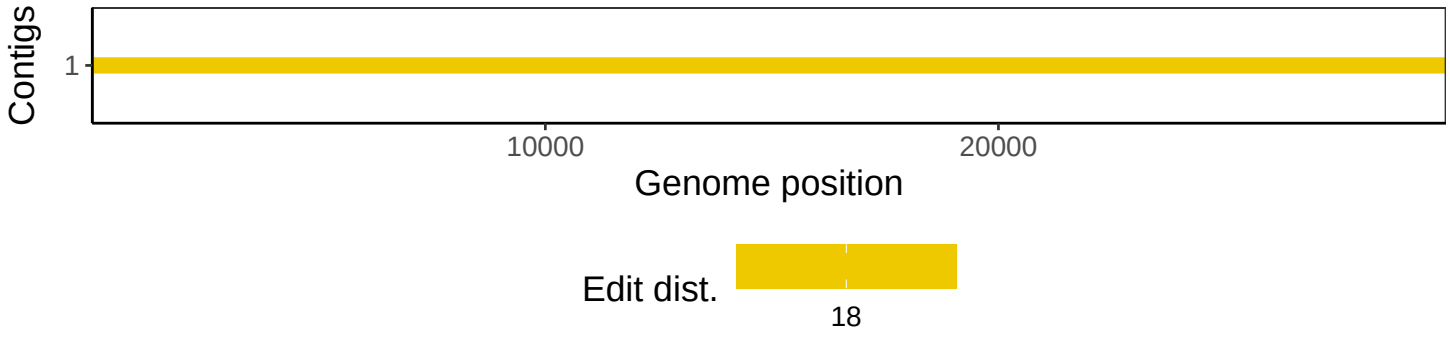
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

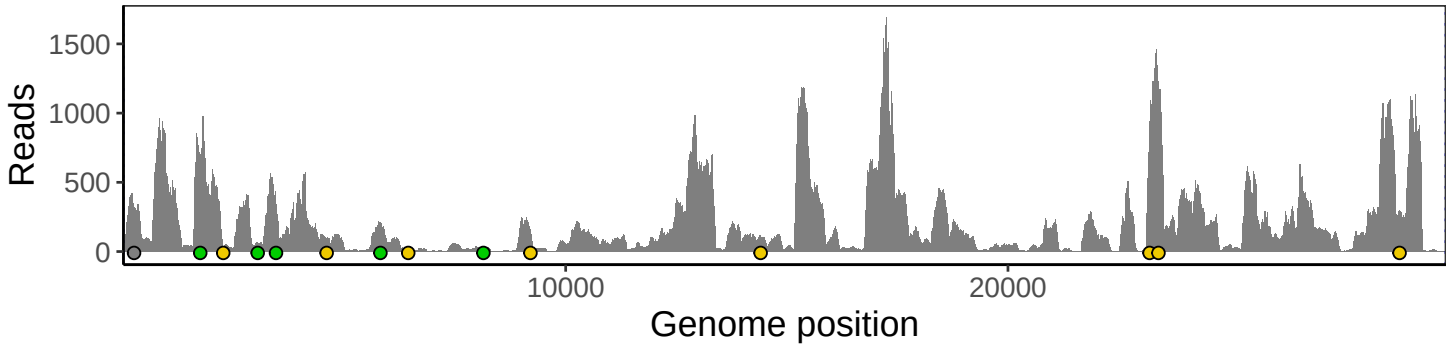


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

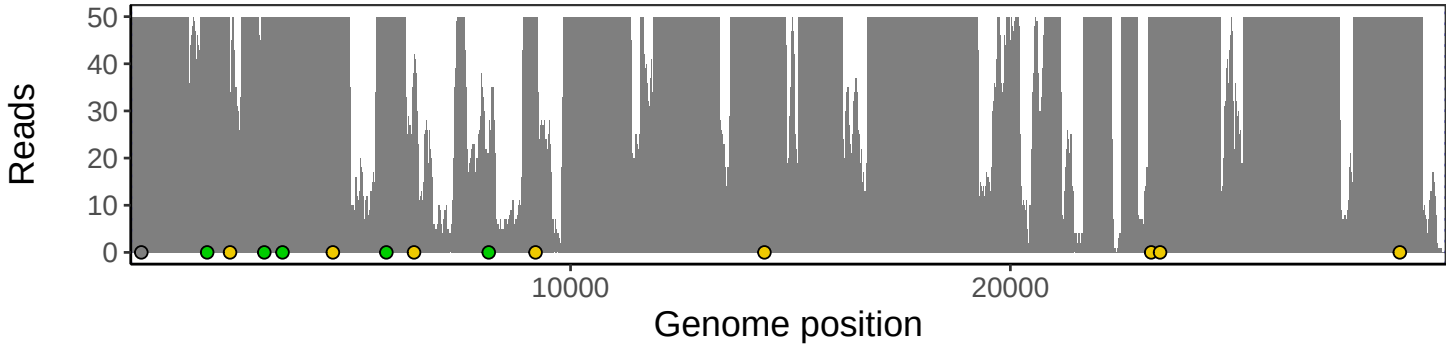


VSP0806-4 | 2021-02-11 | Saline | MPCluster2-Seq14 | genomes | single experiment

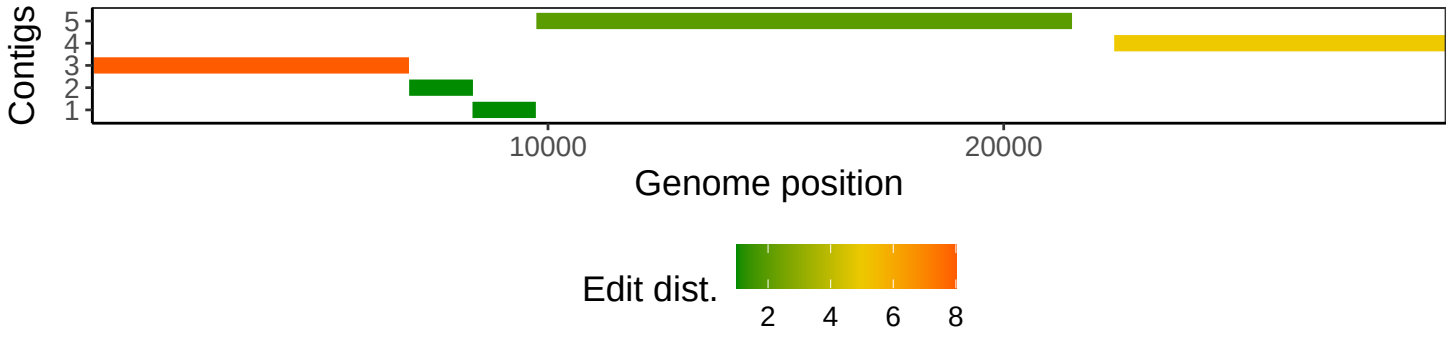
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

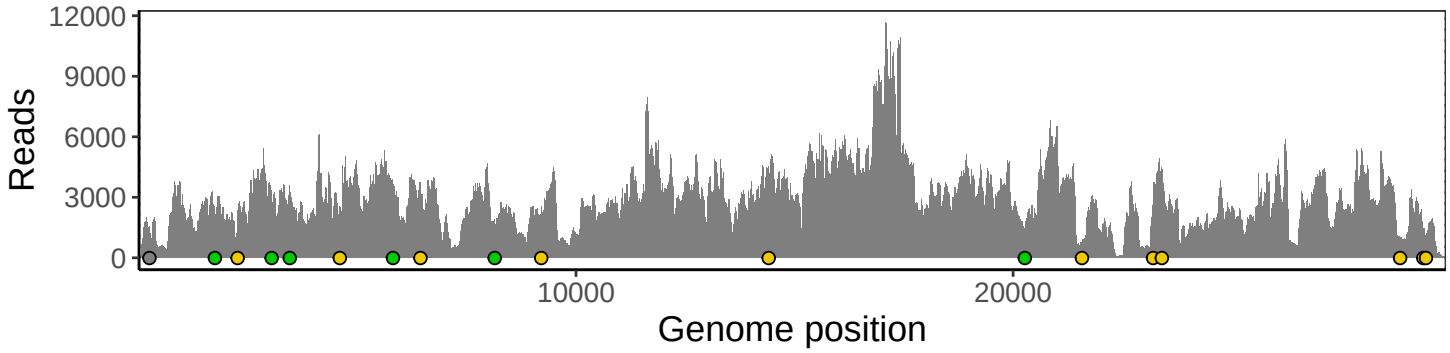


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

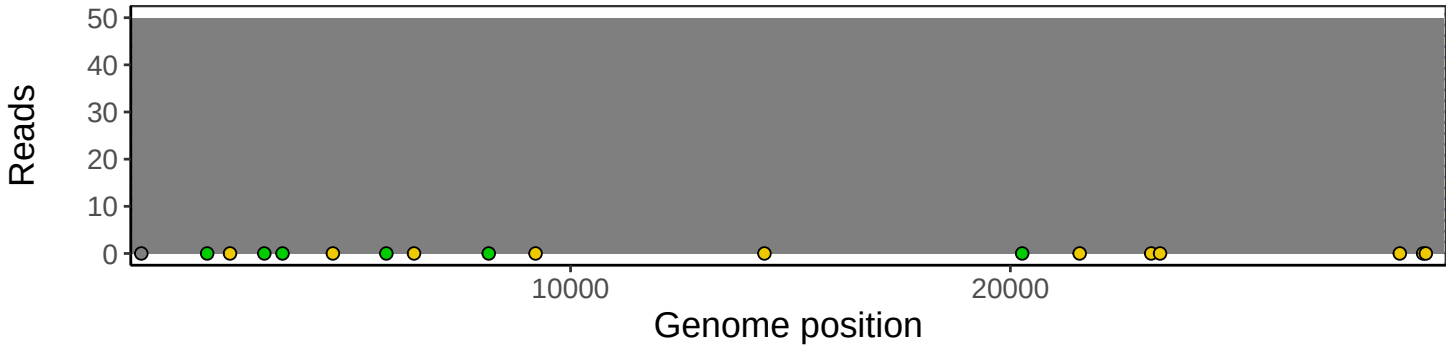


VSP0806-5 | 2021-02-11 | Saline | MPCluster2-Seq14 | genomes | single experiment

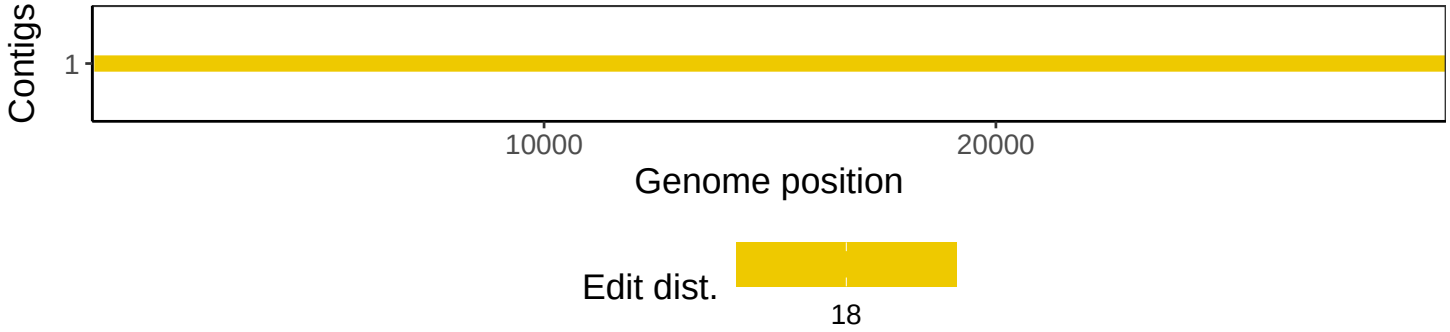
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

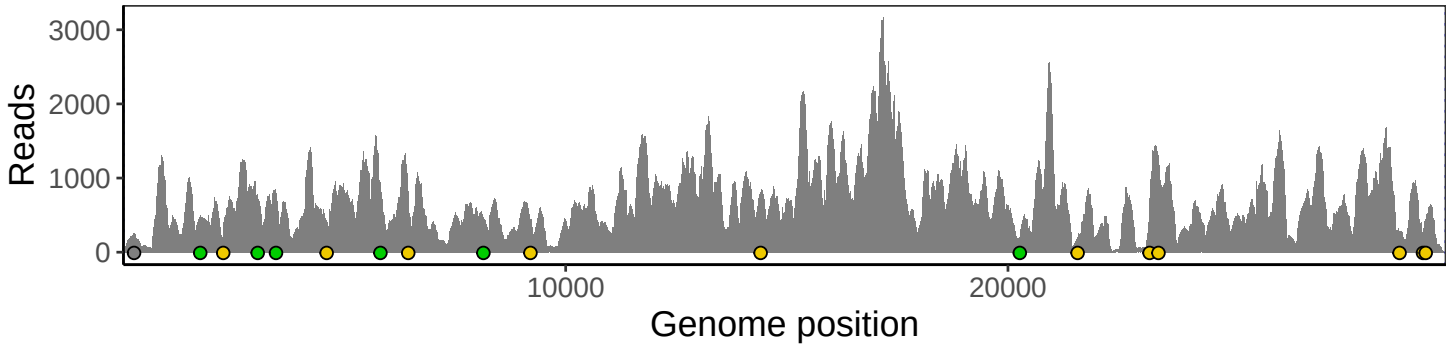


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

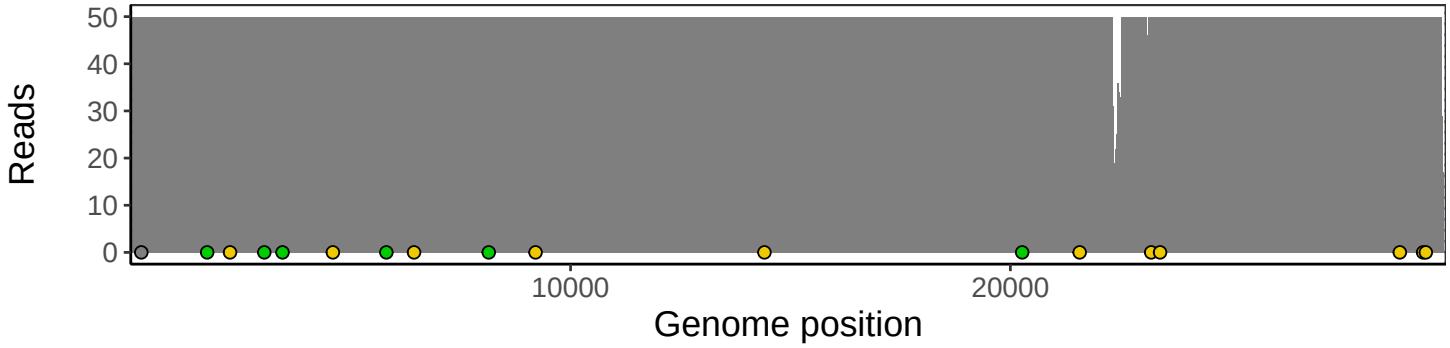


VSP0806-6 | 2021-02-11 | Saline | MPCluster2-Seq14 | genomes | single experiment

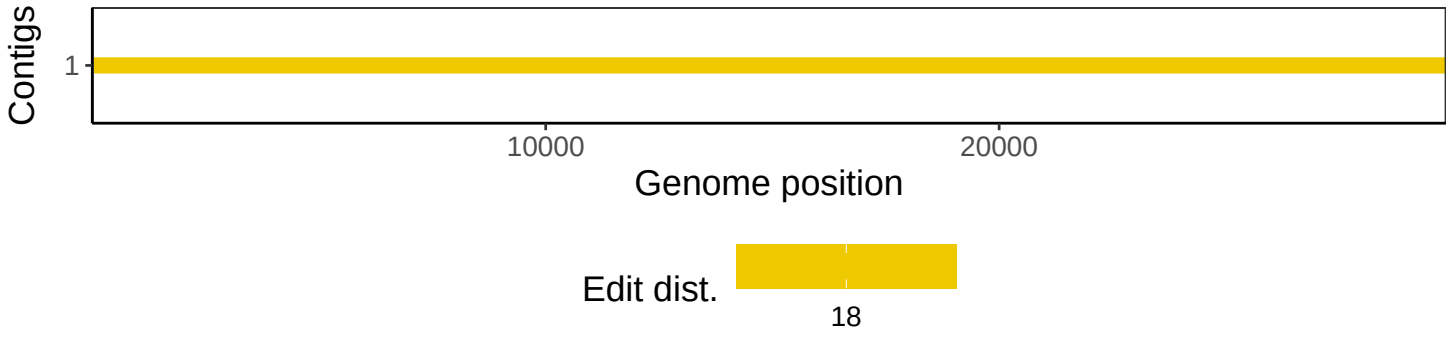
The plot below shows the number of reads covering each nucleotide position in the reference genome. Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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