COVID-19 subject 238

2020-08-18

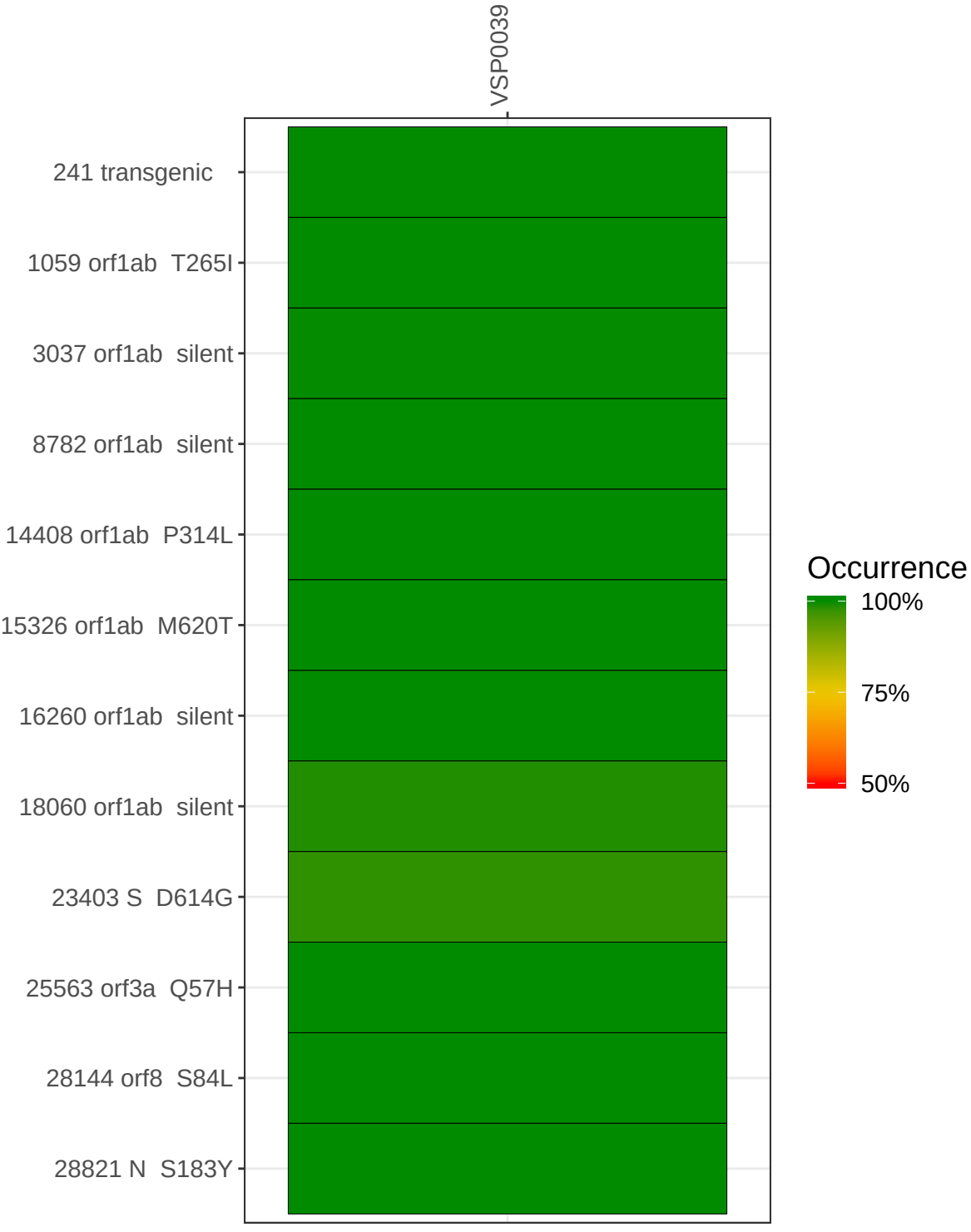
The table below provides a summary of subject samples for which sequencing data is available.
The experiments column shows the number of sequencing experiments performed for each specimen.
Experiment specific analyses are shown at the end of this report.

Table 1. Sample summary.

Experiment	Type	Input genomes	Sample type	Sample date	Largest contig (KD)	Reference read coverage	Reference read coverage (>= 5 reads)
VSP0039	composite	NA	NP	4/29/2020	29.89	99.8%	99.8%
VSP0039-1a	single experiment	493	NP	4/29/2020	0.52	63.6%	60.6%
VSP0039-1b	single experiment	493	NP	4/29/2020	0.77	66.4%	62.1%
VSP0039-2	single experiment	2465	NP	4/29/2020	29.89	99.8%	99.8%

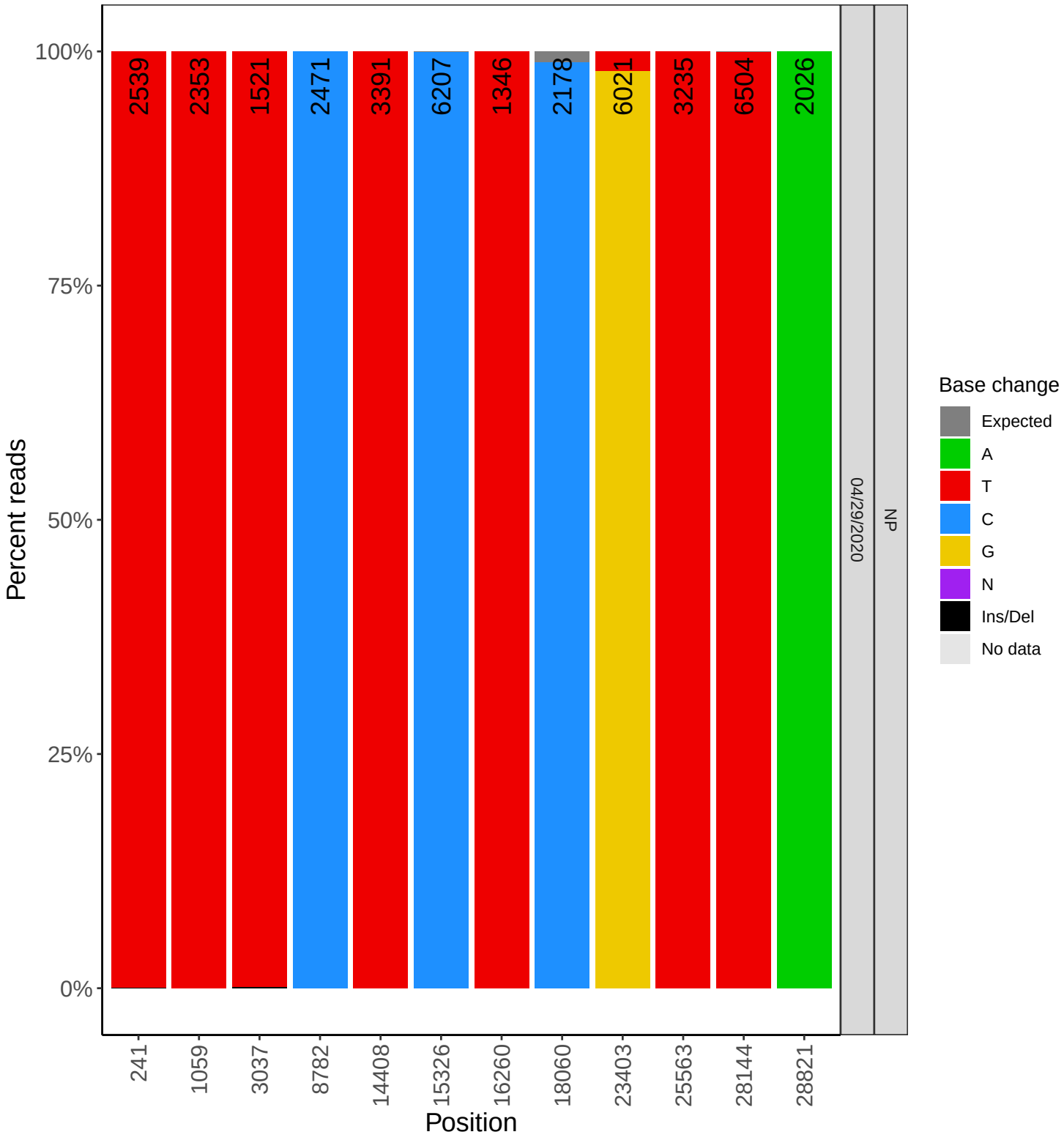
Variants shared across samples

The heat map below shows how variants are shared across subject samples where the percent variance is colored. Variants are called if a variant is covered by 5 for more reads, the alternative base is found in $>$ 50% of reads, the variant and yields a PHRED score $>$ 20 which represents a probabilty of $<$ 1% that a variant is called solely because of sequencing error. Gray tiles denote positions where the variant was not the major variant or no variants were found. The base composition of all tiles are shown in the following plot.



Relative read abundances of variants

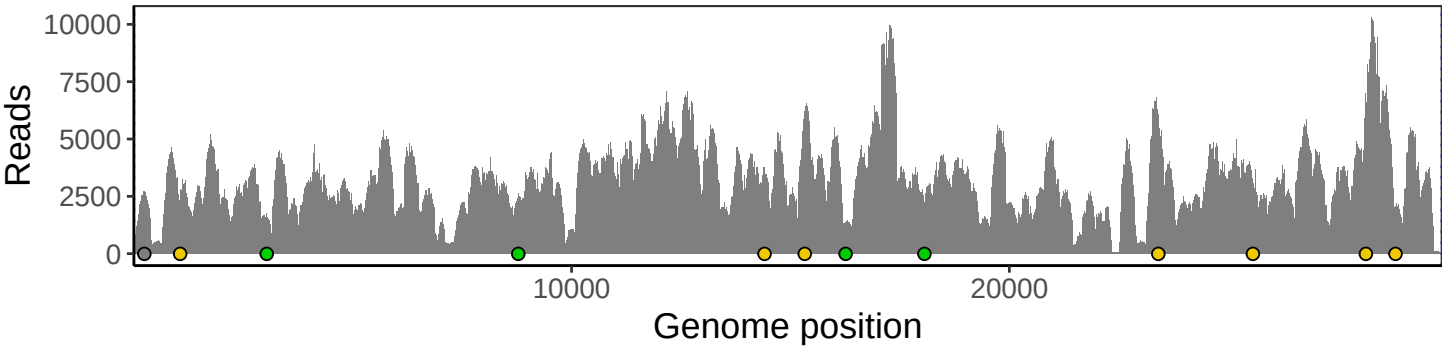
The plot below shows the relative abundances of bases read for each position in the previous variant heatmap where the total number of read pairs covering each position is printed on the stacked bar plots.



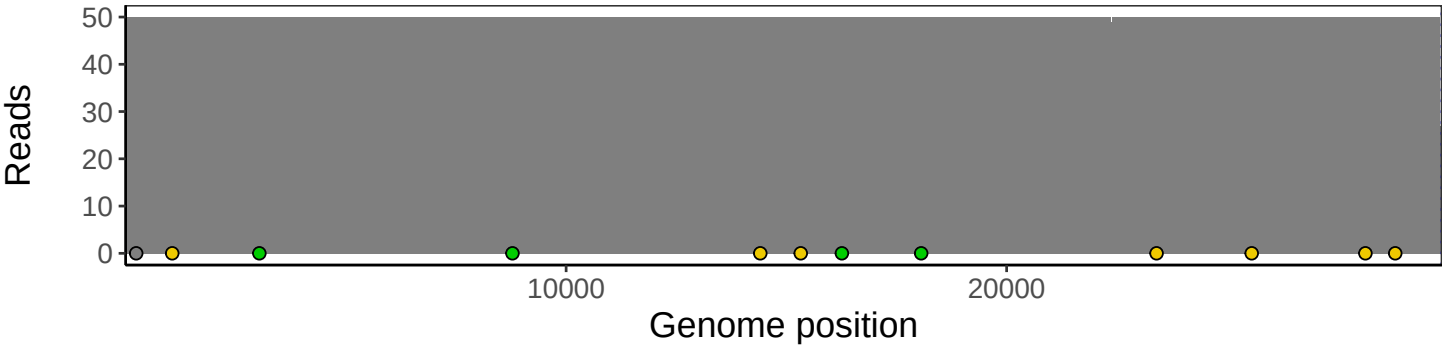
Analyses of individual experiments and composite results.

VSP0039 | 4/29/2020 | NP | 238n-tri | composite result

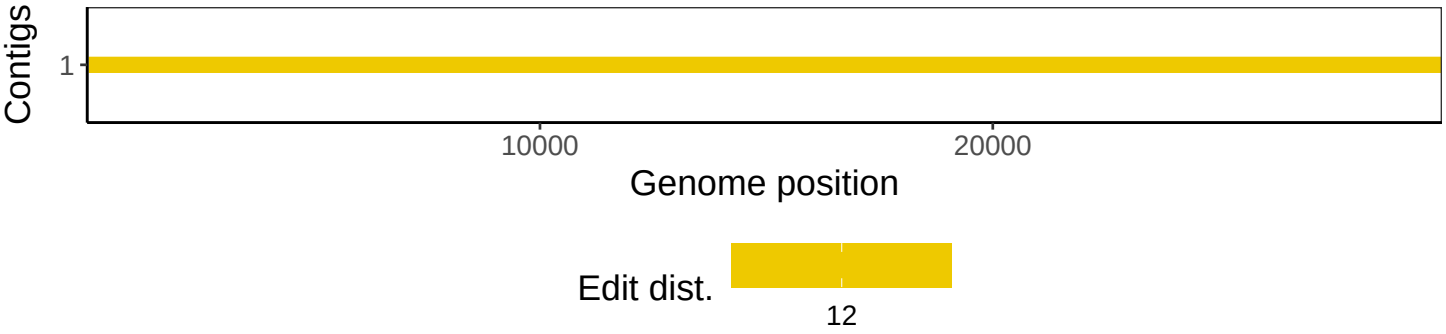
The plot below shows the number of reads covering each nucleotide position in the reference genome (USA-WA1-2020). Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

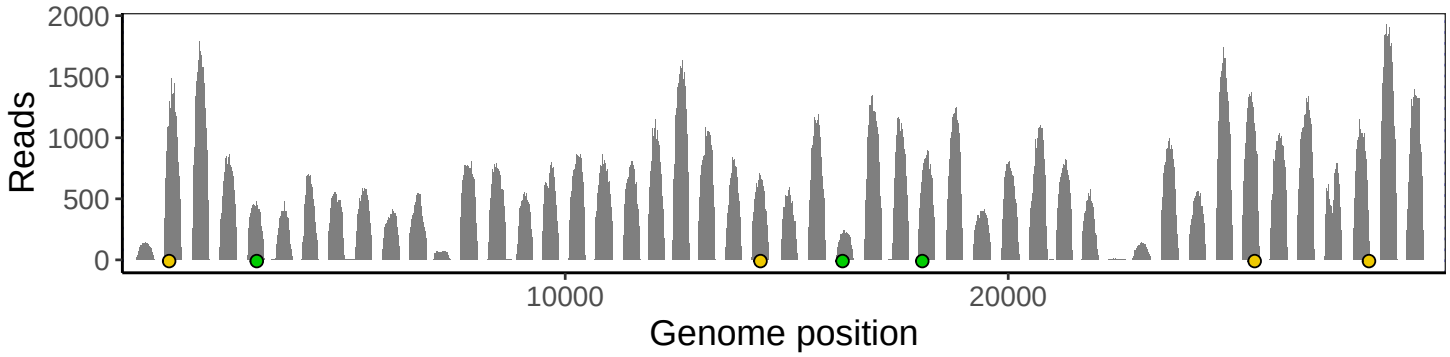


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

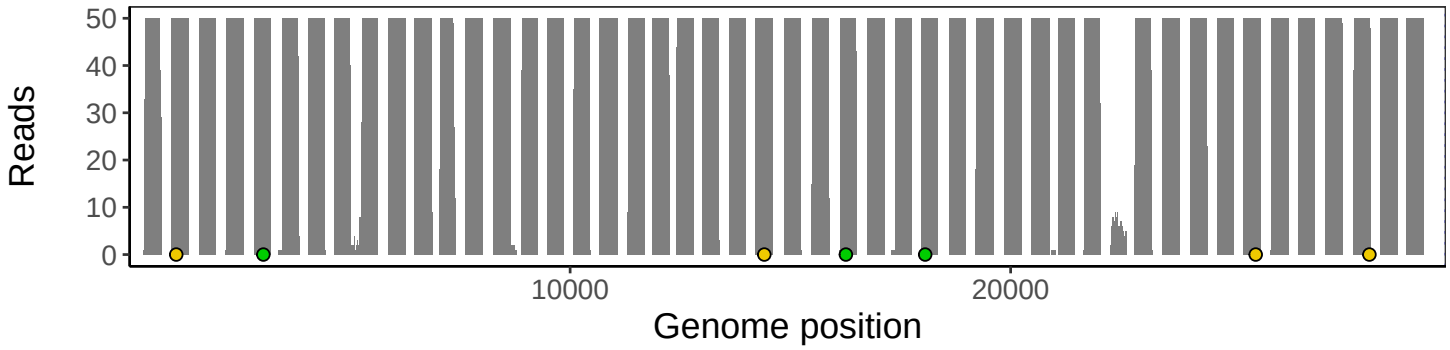


VSP0039-1a | 4/29/2020 | NP | 238n-tri | 493 genomes | single experiment

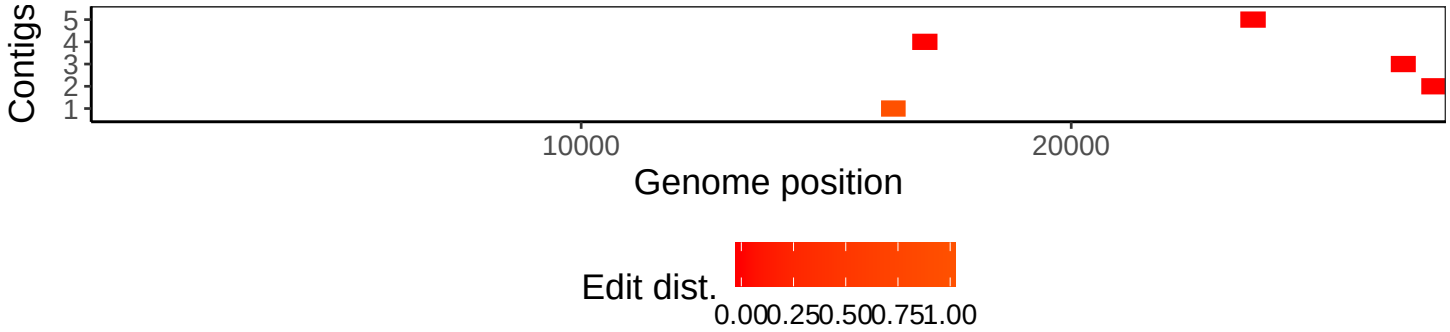
The plot below shows the number of reads covering each nucleotide position in the reference genome (USA-WA1-2020). Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

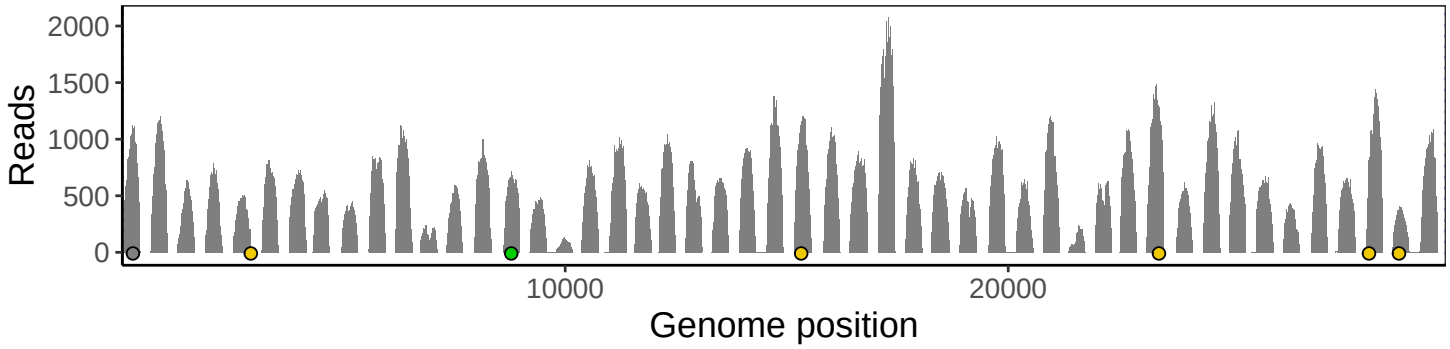


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

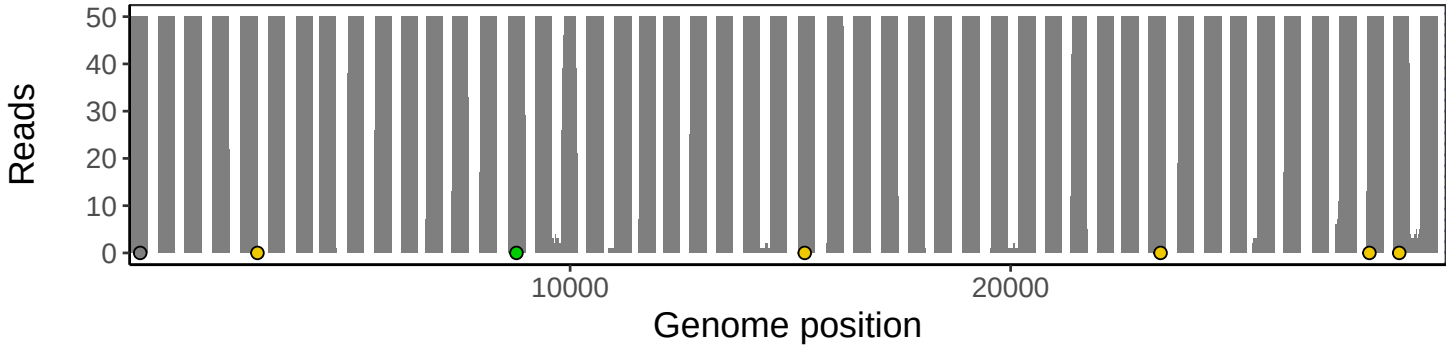


VSP0039-1b | 4/29/2020 | NP | 238n-tri | 493 genomes | single experiment

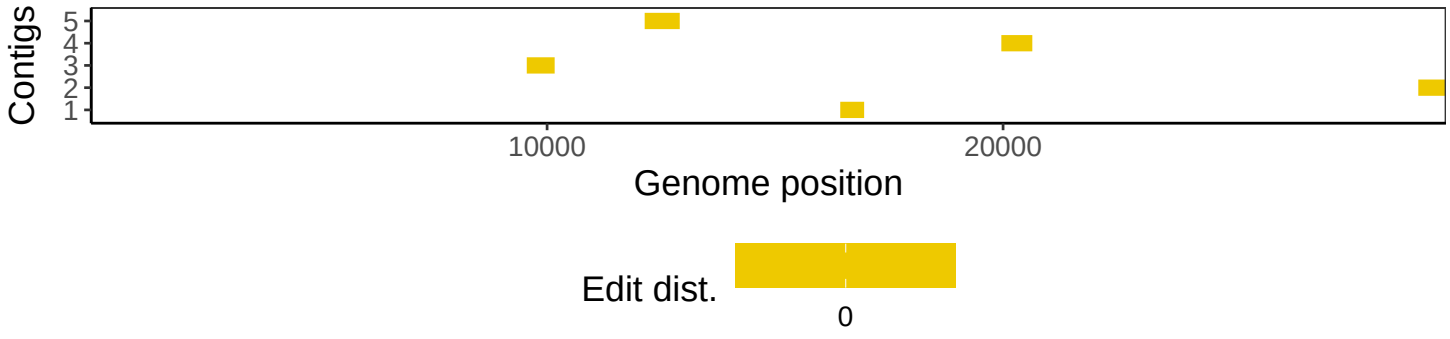
The plot below shows the number of reads covering each nucleotide position in the reference genome (USA-WA1-2020). Variants are shown as colored dots along the bottom of the plot and are color coded according by variant types: gray - transgenic, green - silent, gold - missense, red - nonsense, black - indel.



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

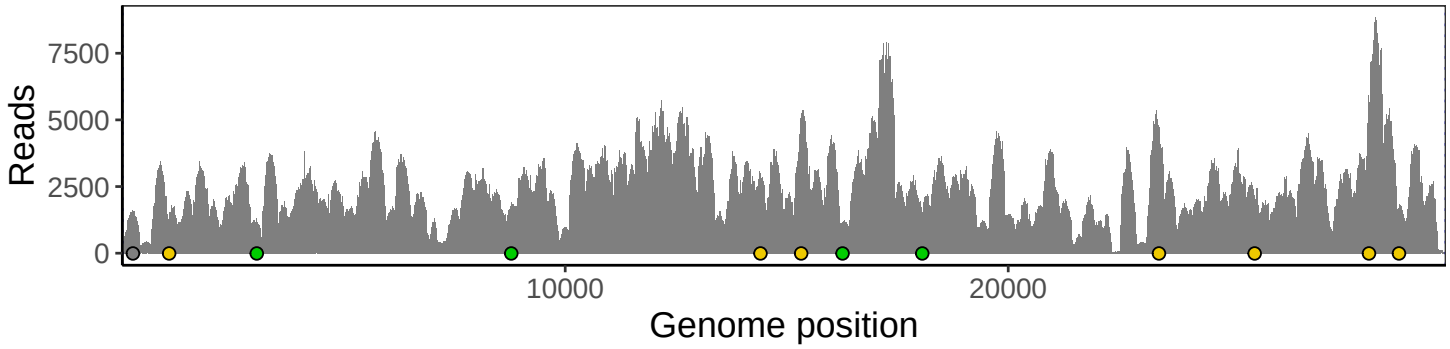


The longest five assembled contigs are shown below colored by their edit distance to the reference genome.

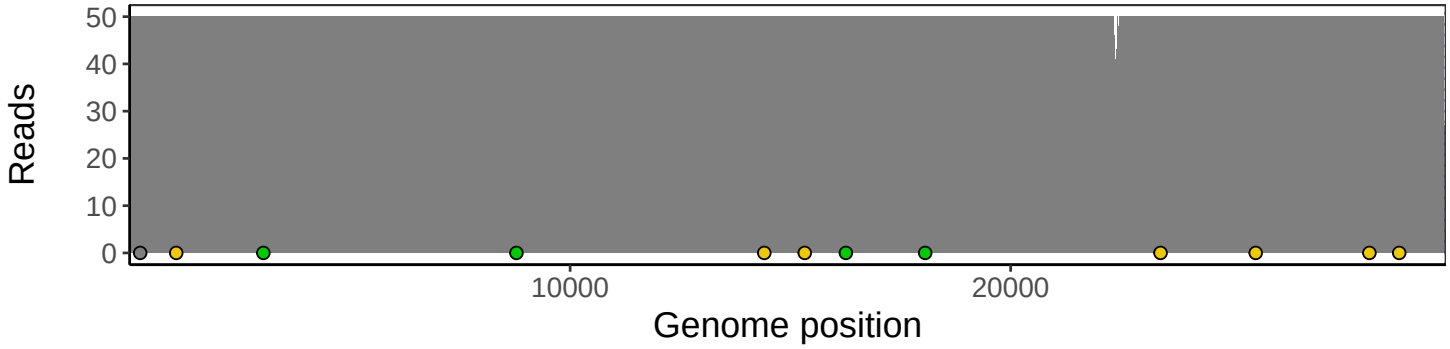


VSP0039-2 | 4/29/2020 | NP | 238n-tri | 2465 genomes | single experiment

The plot below shows the number of reads covering each nucleotide position in the reference genome (USA-WA1-2020). 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.