For a given set of SARS-CoV-2 genomic sites compare predicted probabilities of finding specific codons at given genomic sites (based on the evolutionary analysis of closely related animal sarbecoviruses) vs observed variation with a median of **3078961.5** consensus genomes per codon of SARS-CoV-2 from GISAID.

## Download .JSON data

9440 genomic codon loci analyzed



8688 loci with observed variants



8723 loci with evolutionary predictions



4:8

Median (95%) observed codon variants per locus



4:8

Median (95%) predicted codon variants per locus



0

Median (predicted-observed) count difference per locus



8753

Loci with all variants at ≥ 0.1% predicted



1491

Loci with unpredicted variants at ≥ 0.01%



724

Sites with perfectly predicted minority variants at 0.01% threshold



0.356

Spearkman rank correlation between observed and predicted site entropies



0.274

Fraction of variable loci where the top minority codon was correctly predicted



1.74e+4

Prediction bit-score compared to a matched complexity random model (0.01% threshold)



Decemb	er 2019	November 2020	May 2021	



## Continued evolution, complex selection dynamics, transition to endemic?

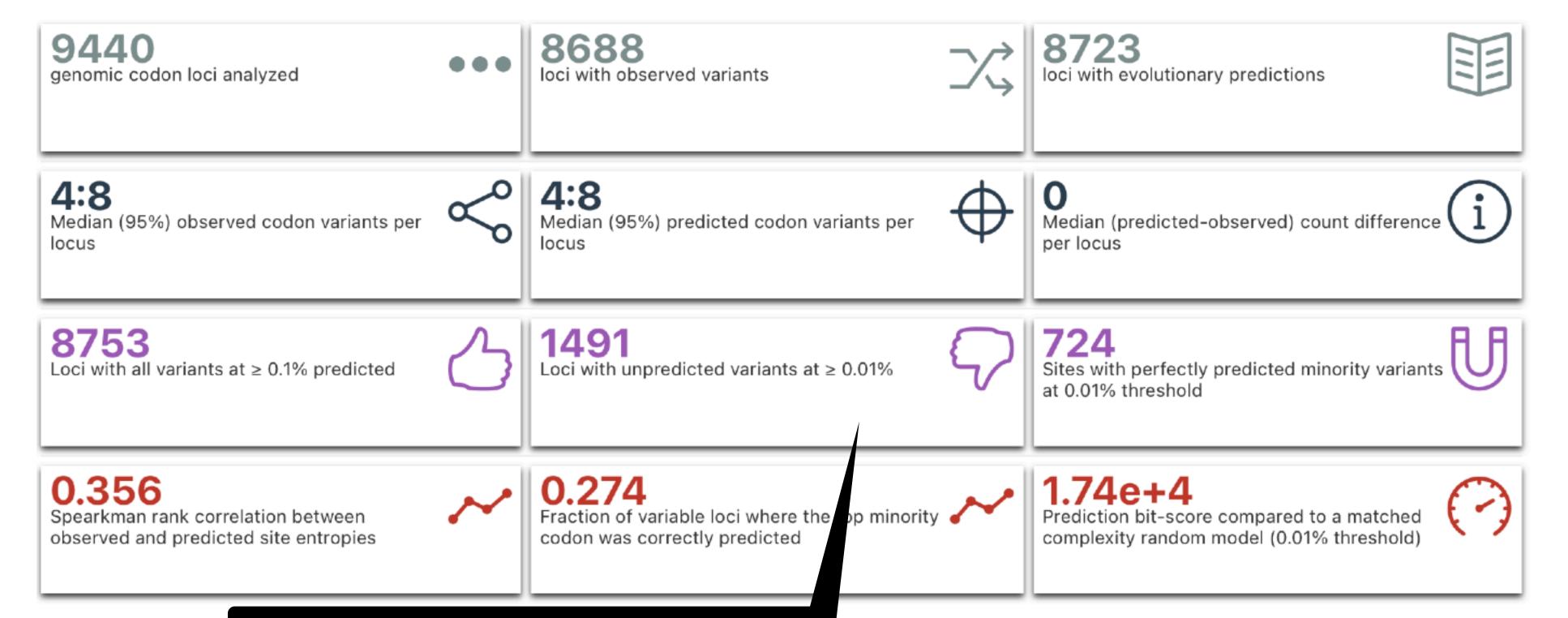
https://observablehq.com/@spond/sars-cov-2-pvo

## The set of "unusual" changes (compared to nCOV)

Continued evolution, complex selection dynamics, transition to endemic?

For a given set of SARS-CoV-2 genomic sites compare predicted probabilities of finding specific codons at given genomic sites (based on the evolutionary analysis of closely related animal sarbecoviruses) vs observed variation with a median of **3078961.5** consensus genomes per codon of SARS-CoV-2 from GISAID.

## Download .JSON data



The set of "unusual" changes (compared to nCOV)

December 2019 November 2020 May 2021

Continued evolution, complex selection dynamics, transition to endemic?