Substitution spectra of SARS-CoV-2 genome reveals insights into the evolution of variants across the pandemic

Javaria Ashraf[1], Ali Raza Bukhari[1], Akbar Kanji[1], Tulaib Iqbal[1], Waqasuddin Khan[2], and Zahra Hasan[1]

[1]The Aga Khan University Hospital
[2]The Aga Khan University Department of Paediatrics and Child Health

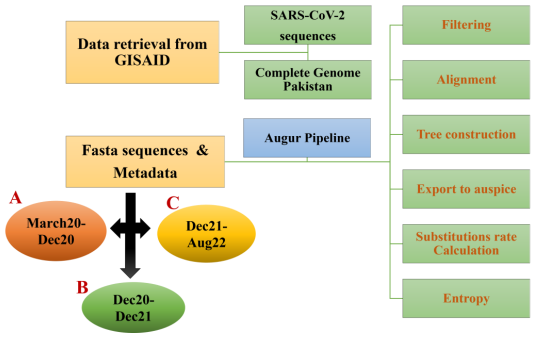
April 4, 2023

Abstract

Background: Changing morbidity and mortality from COVID-19 has been associated with the emergence of new SARS-CoV-2 variants. Whereby, acquisition of mutations in the Spike glycoprotein enhanced host receptor binding, cell entry and antibody escape. Understanding these can help predict the impact of these changes. We used genome sequence data to investigate mutation rates and entropy of SARS-CoV-2 during pandemic surges between 2020 and 2022. **Methods**: 1,637 SARS-CoV-2 genomes from Pakistan were analyzed using the Augur phylogenetic pipeline. Substitution rates and entropy of genomes were calculated year wise and, entropy in the Spike gene was compared for 2020, 2021 and 2022 (defined as periods A, B and C). **Central Findings**: In period A, G clades were predominant and SARS-CoV-2 genome substitution rate was 6.06 x 10^{-4} per site per year. In period B, Delta variant were dominant and substitution rates increased to 9.74 x 10^{-4}. In period C, Omicron variants dominated with substitution rates at 5.02 x 10^{-4}. The rate of genome-wide entropy was the highest during B particularly, in the Spike gene such as, E484K and K417N. During C, genome-wide mutations were increased whilst entropy was reduced. **Conclusions:** The highest SARS-CoV-2 genome substitution rates in 2021 were associated with the Delta wave, which had the greatest morbidity and mortality. These stabilized during the Omicron wave in 2022, when COVID-19 numbers were high mortality was lower. Assessment of SARS-CoV-2 evolution should be monitored together with phylogeographical analysis can help predict future outbreaks and guide public health interventions.

Hosted file

`Mian-Manuscript31-3-23.docx` available at https://authorea.com/users/603101/articles/633606-substitution-spectra-of-sars-cov-2-genome-reveals-insights-into-the-evolution-of-variants-across-the-pandemic



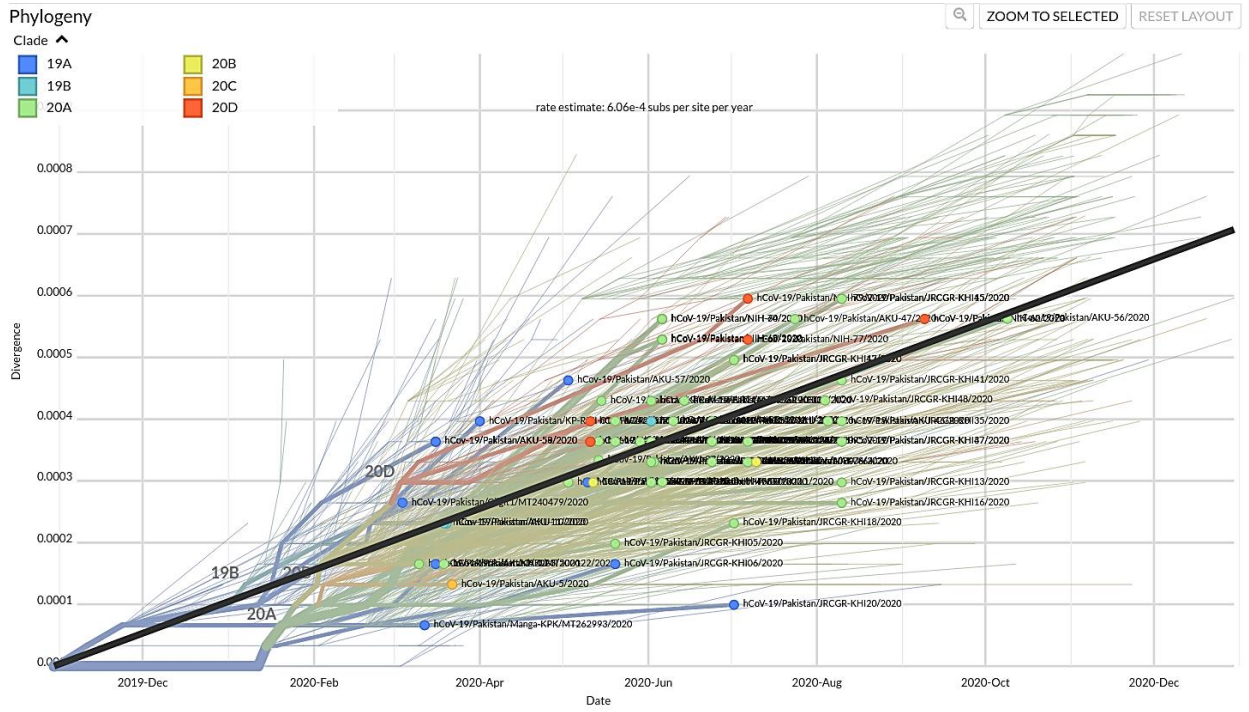
Hosted file

Table1.docx available at https://authorea.com/users/603101/articles/633606-substitution-spectra-of-sars-cov-2-genome-reveals-insights-into-the-evolution-of-variants-across-the-pandemic

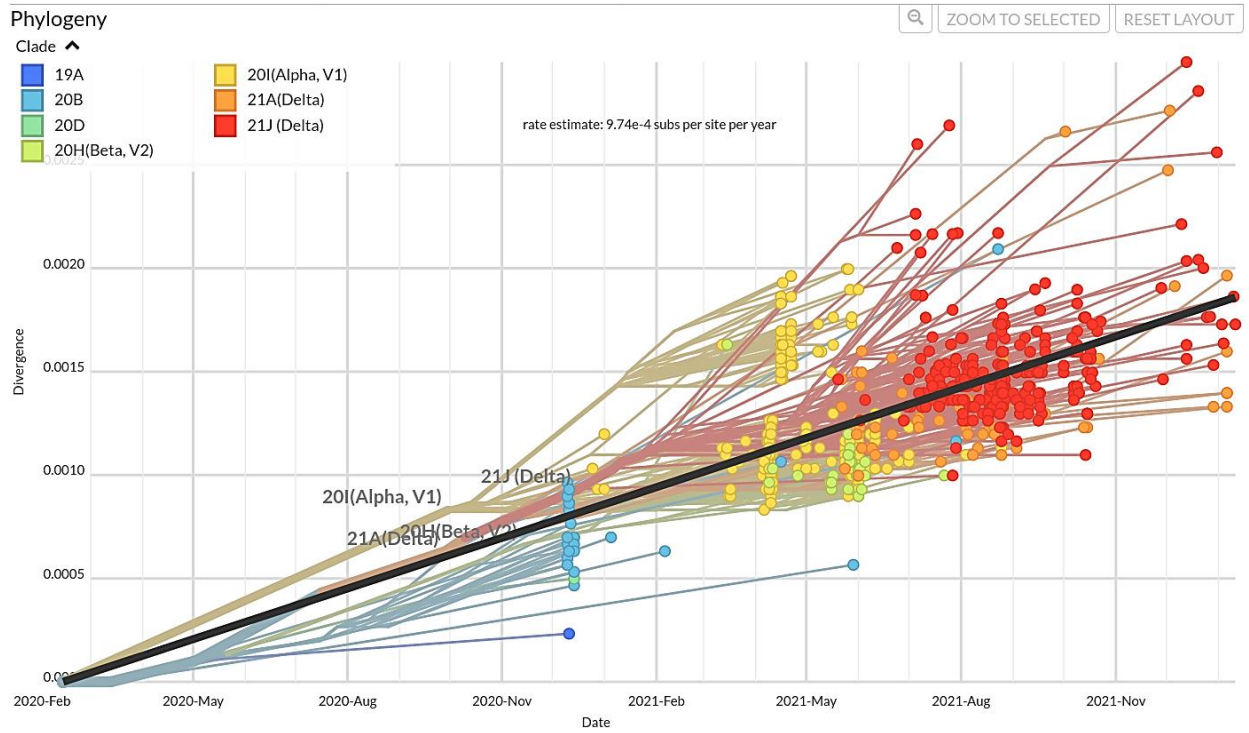
Hosted file

Table2.docx available at https://authorea.com/users/603101/articles/633606-substitution-spectra-of-sars-cov-2-genome-reveals-insights-into-the-evolution-of-variants-across-the-pandemic

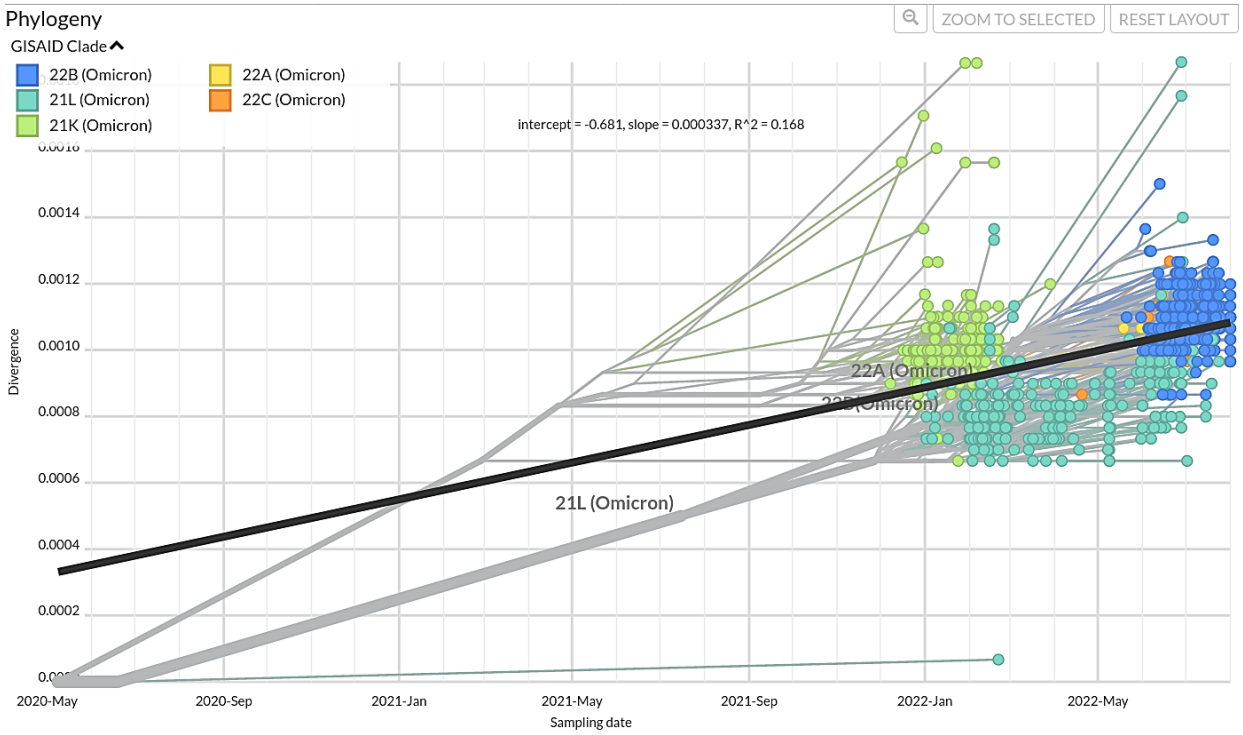
A


Phylogeny
ZOOM TO SELECTED
RESET LAYOUT
Clade
19A
19B
20A
20B
20C
20D
rate estimate: 6.06e-4 subs per site per year
Divergence
0.0008
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
hCoV-19/Pakistan/AKU-57/2020
hCoV-19/Pakistan/AKU-58/2020
hCoV-19/Pakistan/MT240479/2020
hCoV-19/Pakistan/AKU-5/2020
hCoV-19/Pakistan/Manga-KPK/MT262993/2020
hCoV-19/Pakistan/JRCGR-KHI05/2020
hCoV-19/Pakistan/JRCGR-KHI06/2020
hCoV-19/Pakistan/JRCGR-KHI18/2020
hCoV-19/Pakistan/JRCGR-KHI20/2020
hCoV-19/Pakistan/JRCGR-KHI41/2020
hCoV-19/Pakistan/JRCGR-KHI48/2020
hCoV-19/Pakistan/JRCGR-KHI13/2020
hCoV-19/Pakistan/JRCGR-KHI16/2020
hCoV-19/Pakistan/NIH-77/2020
hCoV-19/Pakistan/AKU-56/2020
20D
19B
20A
2019-Dec
2020-Feb
2020-Apr
2020-Jun
2020-Aug
2020-Oct
2020-Dec
Date

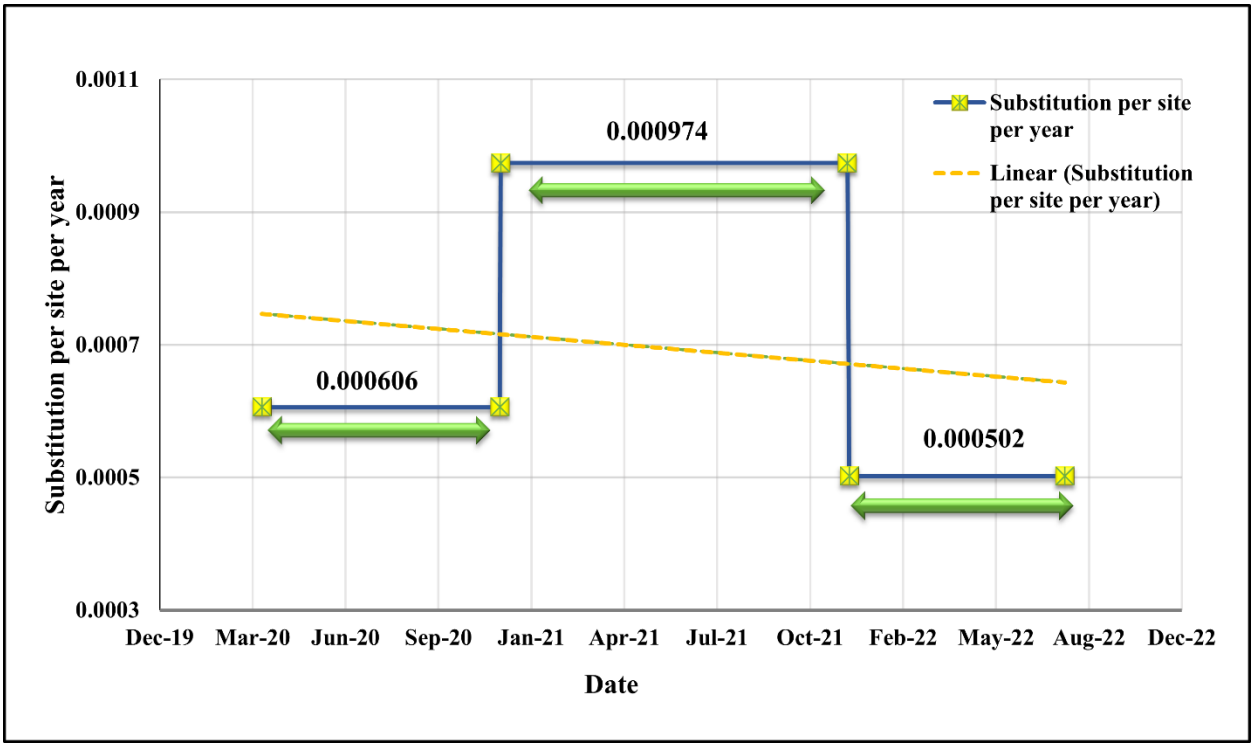
B


Phylogeny
ZOOM TO SELECTED
RESET LAYOUT
Clade
19A
20B
20D
20H(Beta, V2)
20I(Alpha, V1)
21A(Delta)
21J (Delta)
rate estimate: 9.74e-4 subs per site per year
Divergence
0.0020
0.0015
0.0010
0.0005
21J (Delta)
20I(Alpha, V1)
21A(Delta)
20H(Beta, V2)
2020-Feb
2020-May
2020-Aug
2020-Nov
2021-Feb
2021-May
2021-Aug
2021-Nov
Date

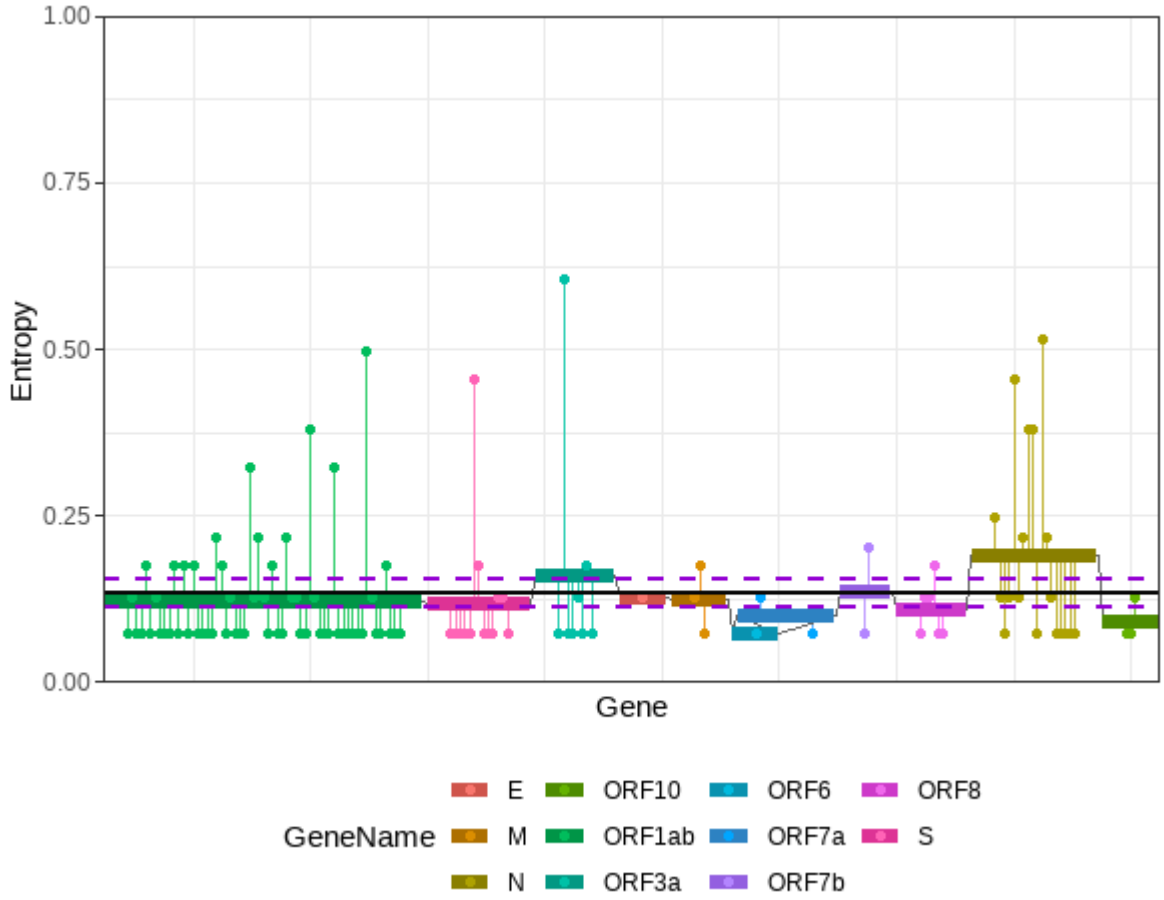
C


Phylogeny
ZOOM TO SELECTED
RESET LAYOUT
GISAID Clade
22B (Omicron)
22A (Omicron)
21L (Omicron)
22C (Omicron)
21K (Omicron)
intercept = -0.681, slope = 0.000337, R^2 = 0.168
Divergence
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
22A (Omicron)
21L (Omicron)
2020-May
2020-Sep
2021-Jan
2021-May
2021-Sep
2022-Jan
2022-May
Sampling date

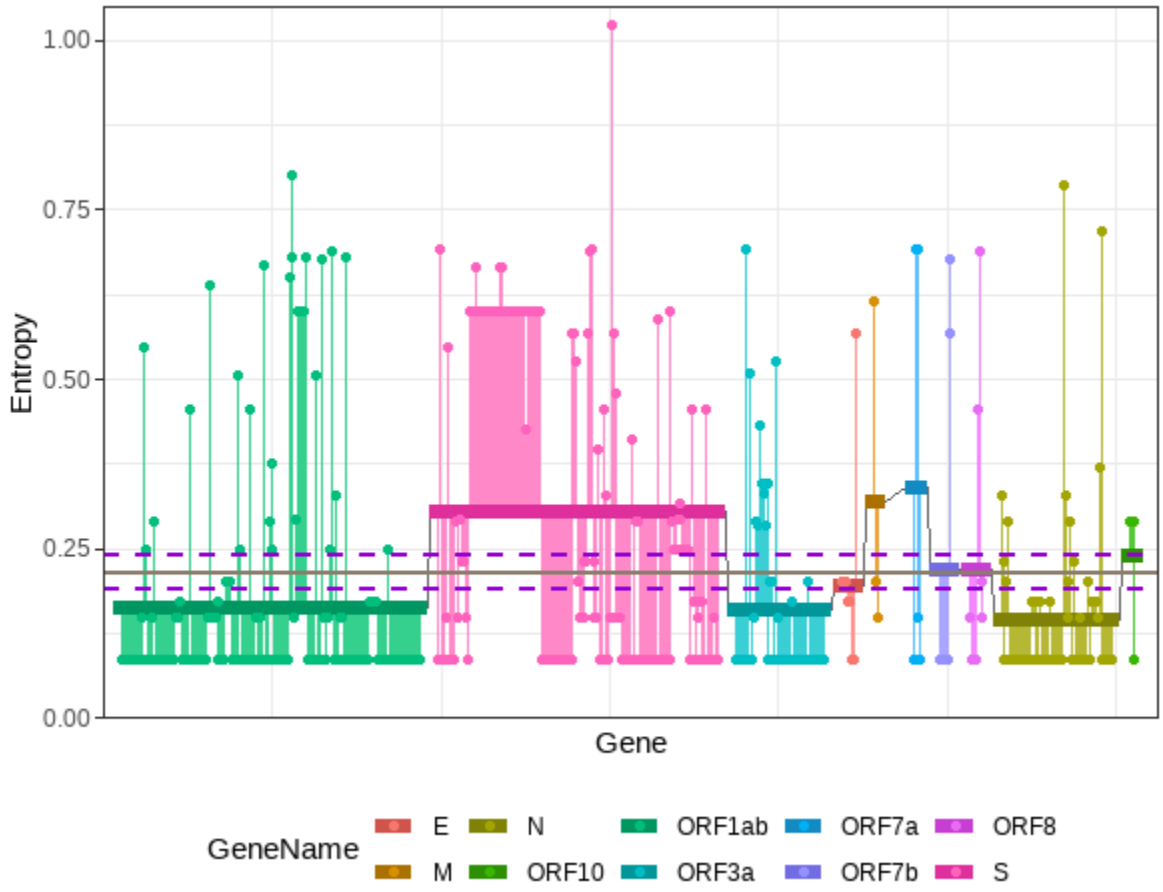
D


0.0011
0.0009
0.0007
0.0005
0.0003
Substitution per site per year
0.000974
0.000606
0.000502
Substitution per site per year
Linear (Substitution per site per year)
Dec-19
Mar-20
Jun-20
Sep-20
Jan-21
Apr-21
Jul-21
Oct-21
Feb-22
May-22
Aug-22
Dec-22
Date

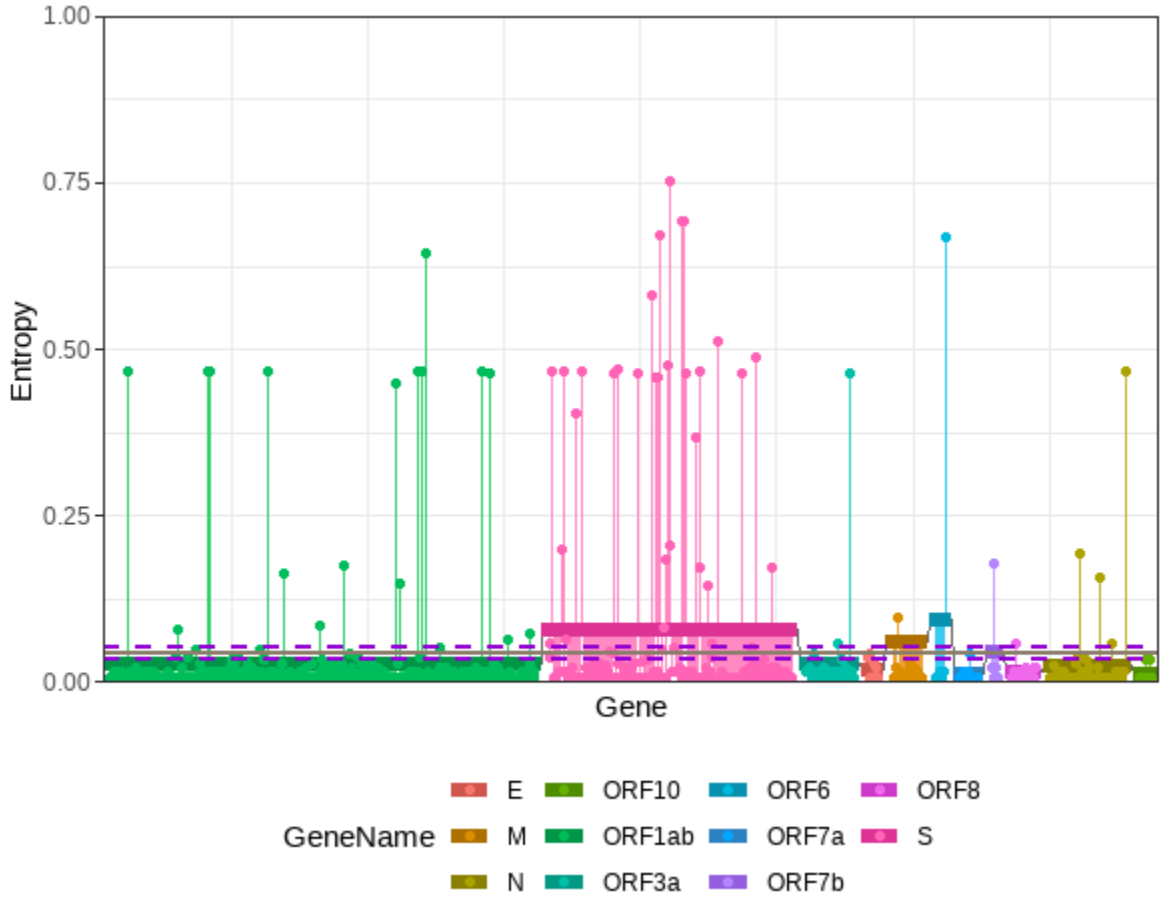
A


1.00
0.75
0.50
0.25
0.00
Entropy
Gene
GeneName
E
M
N
ORF10
ORF1ab
ORF3a
ORF6
ORF7a
ORF7b
ORF8
S

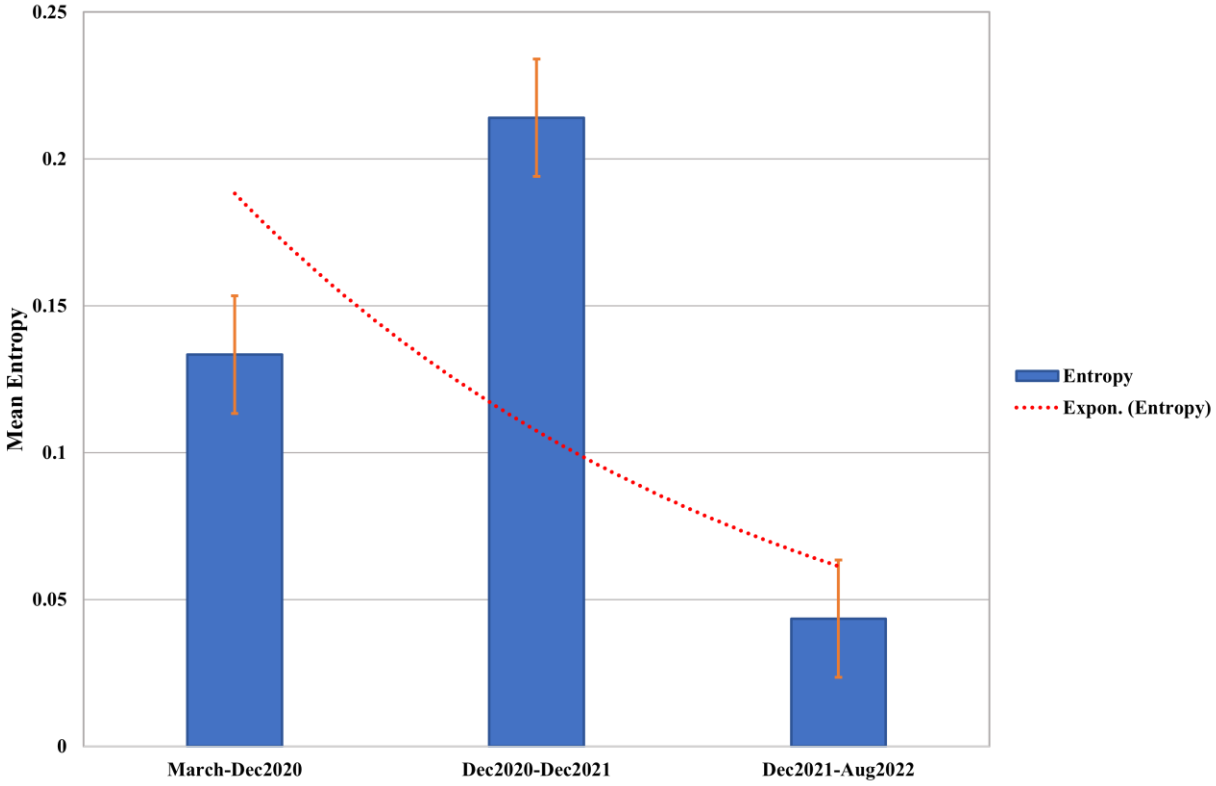
B


1.00
0.75
0.50
0.25
0.00
Entropy
Gene
GeneName
E
M
N
ORF10
ORF1ab
ORF3a
ORF7a
ORF7b
ORF8
S

C


1.00
0.75
0.50
0.25
0.00
Entropy
Gene
GeneName
E
M
N
ORF10
ORF1ab
ORF3a
ORF6
ORF7a
ORF7b
ORF8
S

D


0.25
0.2
0.15
0.1
0.05
0
Mean Entropy
March-Dec2020
Dec2020-Dec2021
Dec2021-Aug2022
Entropy
Expon. (Entropy)