RESEARCH ARTICLE

Prediction of the Effects of Synonymous Variants on SARS-CoV-2 Genome [version 1; peer review: 1 approved with reservations, 2 not approved]

Wan Xin Boon , Boon Zhan Sia, Chong Han Ng

Faculty of Information Science and Technology, Multimedia University, Bukit Beruang, Melaka, 75450, Malaysia

V1 First published: 18 Oct 2021, 10:1053
<https://doi.org/10.12688/f1000research.72896.1>
 Latest published: 05 Sep 2022, 10:1053
<https://doi.org/10.12688/f1000research.72896.2>

Abstract

Background: The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) had led to a global pandemic since December 2019. SARS-CoV-2 is a single-stranded RNA virus, which mutates at a higher rate. Multiple studies had been done to identify and study nonsynonymous mutations, which change amino acid residues of SARS-CoV-2 proteins. On the other hand, there is little study on the effects of SARS-CoV-2 synonymous mutations. Although these mutations do not alter amino acids, some studies suggest that they may affect viral fitness. This study aims to predict the effect of synonymous mutations on the SARS-CoV-2 genome.

Methods: A total of 30,229 SARS-CoV-2 genomic sequences were retrieved from Global Initiative on Sharing all Influenza Data (GISAID) database and aligned using MAFFT. Then, the mutations and their respective frequency were identified. A prediction of RNA secondary structures and their base pair probabilities was performed to study the effect of synonymous mutations on RNA structure and stability. Relative synonymous codon usage (RSCU) analysis was also performed to measure the codon usage bias (CUB) of SARS-CoV-2.

Results: A total of 150 synonymous mutations were identified. The synonymous mutation identified with the highest frequency is C3037U mutation in the nsp3 of ORF1a, followed by C313U and C9286U mutation in nsp1 and nsp4 of ORF1a, respectively.

Conclusion: Among the synonymous mutations identified, C913U mutation in ORF1a and C26735U in membrane (M) protein may affect RNA secondary structure, reducing the stability of RNA folding and possibly resulting in a higher translation rate. However, lab experiments are required to validate the results obtained from prediction analysis.

Keywords

COVID-19, SARS-CoV-2, Synonymous mutations, RNA secondary structure

Open Peer Review

Approval Status

	1	2	3
version 2 (revision) 05 Sep 2022			 view
version 1 18 Oct 2021	 view	 view	 view

1. **Takahiko Koyama** , IBM TJ Watson Research Center, Yorktown Heights, USA
2. **Leyi Wang** , University of Illinois, Urbana, USA
3. **Chandran Nithin** , International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the **Research Synergy Foundation** gateway.

This article is included in the **Emerging Diseases and Outbreaks** gateway.

This article is included in the **Coronavirus** collection.

Corresponding author: Chong Han Ng (chng@mmu.edu.my)

Author roles: Boon WX: Formal Analysis, Investigation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Sia BZ: Investigation, Methodology; Ng CH: Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This research is supported by Multimedia University, Malaysia, IRFund 2.0 (grant number MMUI/210119 awarded to Chong Han, Ng). The funder has no role in study design, data analysis, decision to publish or manuscript preparation. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2021 Boon WX *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Boon WX, Sia BZ and Ng CH. **Prediction of the Effects of Synonymous Variants on SARS-CoV-2 Genome [version 1; peer review: 1 approved with reservations, 2 not approved]** F1000Research 2021, 10:1053 <https://doi.org/10.12688/f1000research.72896.1>

First published: 18 Oct 2021, 10:1053 <https://doi.org/10.12688/f1000research.72896.1>

Introduction

In December 2019, coronavirus disease 2019 (COVID-19) cases first emerged from Wuhan, China¹. Soon after, rapid spread of COVID-19 has resulted in a serious global outbreak. COVID-19 is an infectious and potentially lethal disease caused by a newly found coronavirus strain, known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The virus causes clinical manifestation ranging from asymptomatic to severe pneumonia and eventually death². SARS-CoV-2 seems to have a higher transmission rate³ but lower mortality rate² in comparison to Middle East respiratory coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV).

SARS-CoV-2 is a single-stranded RNA virus with a genome size of 29,903 bases. In general, RNA viruses have a higher mutation rate than DNA viruses and this allows them to evolve rapidly, escaping the host immune defence response⁴. A study by Kim *et al.* (2020) identified a total of 1,352 nonsynonymous and 767 synonymous mutations from 4,254 SARS-CoV-2 genomes⁵. While in another study done by Khailany *et al.* (2020), 116 mutations had been identified from 95 SARS-CoV-2 genomes⁶.

In this study, we focus on synonymous mutations instead of non-synonymous mutations as researchers often overlook their biological importance. Synonymous mutations are also known as silent mutations because the nucleotide mutations result in a change in the RNA sequence without altering the amino acid sequence⁷. Synonymous mutations have been suggested to have no functional consequence on the fitness of organisms and their evolution in long term⁸. However, numerous recent studies had showed that synonymous mutations may affect the folding and stability of RNA structures⁹. For RNA viruses, even though synonymous mutations generally do not change their pathogenicity, some studies reveal that synonymous mutations may affect the RNA secondary structure of the virus¹⁰ and also change the codon usage bias of the genes in the virus^{11,12}.

Methods

Sequence retrieval

30,229 SARS-CoV-2 genomic sequences were retrieved from GISAID database (Global Initiative on Sharing All Influenza Data, RRID:SCR_018251)¹³ ranging from 31 December 2019 to 22 March 2021. SARS-CoV-2 genomic sequences were filtered by setting parameters to keep only sequences with complete genome and high coverage. The reference sequence of SARS-CoV-2 genome (NC_045512.2)¹⁴ was retrieved in fasta format from NCBI database (NCBI, RRID:SCR_006472). It is a Wuhan isolate with a complete genome which comprises of 29,903 bases.

Multiple sequence alignment

The rapid calculation available in MAFFT online server (MAFFT, version 7.467, RRID:SCR_011811)¹⁵ was used to perform multiple sequence alignment (MSA) for 30,229 SARS-CoV-2 genomes. This option supports the alignment of more than 20,000 sequences with approximately 30,000 sites. The alignment length was kept, which means the insertions at the

mutated sequences were removed, to keep the alignment length the same as the reference sequence. While other parameters were left as default.

Identification of mutations and their frequency in SARS-CoV-2 genomes

A simple Python script was written to identify the mutations in 30,229 SARS-CoV-2 genomes. To determine whether the identified mutations are synonymous or nonsynonymous, MEGA X software, version 10.2.5 build 10210330 (MEGA Software, RRID:SCR_000667)¹⁶ was utilized to perform the translation for inspection purposes. The presence of amino acid changes was identified by referring to the genomic position of the nucleotide mutations. Synonymous mutations with the top 10 highest frequencies were generated.

SARS-CoV-2 RNA secondary structure prediction

The RNA secondary structure of wild type and mutant sequences were predicted using RNAfold program, version 2.4.18 (Vienna RNA, RRID:SCR_008550) to show how mutations affect RNA secondary structure. The RNA secondary structure prediction was performed using a sequence length of 250 nucleotides upstream and downstream of the mutation site. The minimum free energy (MFE) was also calculated in the RNAfold program for both wild type and mutant sequences to show comparison of the RNA folding stability between them. The comparison on the value of minimum free energy (MFE) is important to indicate whether the mutations affect the folding stability of the respective RNA structure.

Base pair probability estimation and analysis

To predict how the mutations affect RNA local folding, base pair probability was estimated by utilizing MutaRNA, version 1.3.0 (MutaRNA, RRID:SCR_021723)¹⁷. MutaRNA is a web-based tool that allows prediction and visualization of the structure changes induced by a single nucleotide polymorphism (SNP) in an RNA sequence. It includes the base pair probabilities within RNA molecule of both wild type and mutant. The parameters used in MutaRNA were set as default, in which the window size is 200nt and the maximal base pair span is 150nt.

Relative synonymous codon usage (RSCU)

Relative synonymous codon usage (RSCU) represents the ratio of the observed frequency of codons appearing in a gene to the expected frequency under equal codon usage. RSCU is calculated using the formula:

$$RSCU_i = \frac{X_i}{\frac{1}{n} \sum_{i=1}^n X_i},$$

where X_i implies the number of occurrences of codon i and n stands for the number of synonymous codons encoded for that particular amino acid.

Results and discussion

A synonymous mutation is a change in the nucleotide that does not cause any changes in the encoded amino acid. Synonymous

mutations were previously considered to be less important, but they are now proven to have some effects on RNA folding, RNA stability, miRNA binding and translational efficiency¹⁸. Synonymous mutations may have significant effects on the adaptation, virulence, and evolution of RNA viruses¹⁹. Another study done also indicated that synonymous mutations have association with more than 50 human diseases such as hemophilia B, tuberculosis (TB), cystic fibrosis (CF), Alzheimer, schizophrenia, chronic hepatitis C and so on²⁰. All these studies show that increasing importance has been associated with synonymous mutations over these years. Hence, it is necessary for us to study the effects of synonymous mutations of SARS-CoV-2 genome.

Identification of SARS-CoV-2 synonymous mutations

A total of 381 mutations were found in SARS-CoV-2 genomes by using python script, in which 150 of them are synonymous mutations. The distribution of synonymous mutations in 11 coding regions is shown in [Figure 1](#). Among these mutations, ORF1a and ORF1b have a higher number of synonymous mutations at 76 and 33, respectively, which might be due to their longer sequence length. Besides that, our findings also show high C to U mutation rate in SARS-CoV-2 genetic variation and this mutational skews are in line with the studies done by Rice *et al.* (2021) and Simmonds (2020)^{21,22}. These mutational skews are necessary to be considered when deducing the selection acting on synonymous variants in SARS-CoV-2 evolution²³.

The synonymous mutations in SARS-CoV-2 genomes with the top 10 highest frequency were listed in [Table 1](#). As shown in [Table 1](#), synonymous mutations with the highest frequency identified from SARS-CoV-2 genomes is C3037U mutation

located in nsp3 of ORF1a, followed by C313U mutation in nsp1 of ORF1a and C9286U mutation in nsp4 of ORF1a. Mutations with higher frequency are mostly found in ORF1a and ORF1b. It is of great interest to find out the effect of these top 10 synonymous mutations on SARS-CoV-2 genome.

RNA secondary structure and base pair probability

SARS-CoV-2 virus can form highly structured RNA elements, which may affect viral replication, discontinuous transcription and translation^{24–28}

The RNA secondary structures of wild type and mutant sequences were predicted using RNAfold program. Minimum free energy (MFE) of each structure was also calculated to show the folding stability of the respective RNA structure. Among all, C913U mutation and C26735U mutation were found to have a more obvious effect on the predicted RNA secondary structure compared to the wild type.

As shown in [Figure 2\(A\)](#), a single hairpin is formed around the mutant with C913U mutation instead of a multiloop in the wild type. The minimum free energy value of the mutant (- 146.90 kcal/mol) is slightly less negative than that of the wild type (- 147.40 kcal/mol), which makes it a less thermodynamically stable structure compared to the wild type.

To visualize the differences in the base pairing potential induced by the mutation, base pairing probability was estimated using MutaRNA. The circular plots in [Figure 2\(B\)](#) show the base pairing probabilities of both wild type and mutant. As shown in the circular plots, C913U mutation decreased the Watson–Crick base pairing probability near the mutation site, which led to a less stable predicted RNA secondary structure.

Figure 1. Distribution of SARS-CoV-2 synonymous mutations in 11 coding regions.

Table 1. SARS-CoV-2 synonymous mutations with the top 10 highest frequency.

	ORF	Position	Nucleotide Variation	Frequency
1a	nsp1	313	C > U	8569
	nsp2	913	C > U	3213
	nsp3	3037	C > U	28054
		5986	C > U	3261
	nsp4	9286	C > U	4030
1b	nsp12	14676	C > U	3262
		15279	C > U	3244
		16176	U > C	3236
	nsp14	18877	C > U	2944
	M	26735	C > U	2701

C913U mutation is found in the nsp2 of ORF1a in SARS-CoV-2 genome. Nsp2 in SARS-CoV interacts with two human host proteins specifically, which are prohibitin 1 (PHB1) and prohibitin 2 (PHB2)²⁹. This interaction may disrupt the intra-cellular host signalling during the viral infections, rather than taking part in the viral replication²⁹. It is yet to see if nsp2 of SARS-CoV-2 shares same or similar function as that of SARS-CoV.

C26735U mutation is located at the membrane (M) protein. M protein may suppress host immune responses via the interference of Type I interferon production³⁰. C26735U mutation induces changes in the predicted RNA secondary structure by forming an extra multibranch loop at the mutation site as shown in Figure 3(A). The RNA secondary structure formed by the mutant (-134.50 kcal/mol) has a less negative minimum free energy value compared to the wild type (-136.30 kcal/mol), which makes it a less stable structure. While for the circular plots in Figure 3(B), C26735U mutation decreases the base pairing probabilities in flanking regions, which again predicted that the mutation decreases the folding stability of the RNA secondary structure.

Figure 2. The effect of C913U mutation on RNA secondary structure of nsp2 in ORF1a. (A) RNA secondary structure of wild type C913 and mutant U913. **(B)** Circular plots showing the base pairing probabilities with darker shades indicating higher base pairing probabilities and a red arrow pointing to the mutation site.

In short, both C913U and C26735U mutations cause a more drastic change in RNA secondary structure. Given the

Figure 3. The effect of C26735U mutation on RNA secondary structure of M protein. (A) RNA secondary structure of wild type C26735 and mutant U26735. **(B)** Circular plots showing the base pairing probabilities with darker shades indicating higher base pairing probabilities and a red arrow pointing to the mutation site.

shortcomings of prediction tools, it is necessary to check if the changes in RNA secondary structure affect the pathogenicity of SARS-CoV-2 using experimental approaches. Besides that, these two mutations also reduce the folding stability of the RNA secondary structure, which then affects the polypeptide translation and folding. There is evidence suggesting that stable RNA structures play a key role in reducing the translation speed to prevent “ribosomal traffic jams” so that the newly translated polypeptides can fold properly³¹. Hence, both C913U and C26735U mutations increase the translation speed of SARS-CoV-2 RNA but they might cause the nascent polypeptide folding more prone to error during translation.

RSCU analysis of SARS-CoV-2

Codon usage bias (CUB), which is non-random usage of synonymous codons, is common in all species. It is a phenomenon where some codons are preferred over others for a specific amino acid. SARS-CoV-2 replicates using host cell's machinery and synthesizes its protein by utilizing host cellular components. Hence, codon usage bias may affect the replication of viruses³².

Relative synonymous codon usage (RSCU) is a widely used statistical approach³³ that can be used to measure codon usage bias in coding sequences. The RSCU values of SARS-CoV-2 are shown in Table 2 and the most preferred codons for each amino acid are marked in bold. Stop codons (UAA, UAG, UGA)

Table 2. RSCU values of SARS-CoV-2 genome.

Amino Acid	Synonymous Codons	RSCU
Ala	GCA	1.09
	GCC	0.58
	GCG	0.16
	GCU	2.17
Arg	AGA	2.67
	AGG	0.81
	CGA	0.29
	CGC	0.58
	CGG	0.19
	CGU	1.46
Asn	AAC	0.65
	AAU	1.35
Asp	GAC	0.72
	GAU	1.28
Cys	UGC	0.45
	UGU	1.55
Gln	CAA	1.39
	CAG	0.61

Amino Acid	Synonymous Codons	RSCU
Glu	GAA	1.44
	GAG	0.56
Gly	GGA	0.82
	GGC	0.71
	GGG	0.12
	GGU	2.34
His	CAC	0.61
	CAU	1.39
Ile	AUA	0.92
	AUC	0.56
	AUU	1.53
Leu	CUA	0.66
	CUC	0.59
	CUG	0.30
	CUU	1.75
	UUA	1.63
	UUG	1.06
Lys	AAA	1.31
	AAG	0.69
Phe	UUC	0.59
	UUU	1.41
Pro	CCA	1.59
	CCC	0.29
	CCG	0.17
	CCU	1.94
Ser	AGC	0.36
	AGU	1.43
	UCA	1.67
	UCC	0.46
	UCG	0.11
Thr	UCU	1.96
	ACA	1.64
	ACC	0.38
	ACG	0.20
Tyr	ACU	1.78
	UAC	0.78
Val	UAU	1.22
	GUA	0.91
	GUC	0.56
	GUG	0.58
	GUU	1.95

and codons which code for an amino acid uniquely (AUG, UGG) are excluded from RSCU analysis.

Based on the RSCU values, the synonymous codons can be classified into five groups: i) codons with RSCU value equals to 1.0 are unbiased codons; ii) codons with RSCU value > 1.0 are codons preferred in a genome; iii) codons with RSCU value < 1.0 are codons less preferred in a genome; iv) codons with RSCU value > 1.6 are codons which are over-represented in a genome; v) codons with RSCU value < 1.6 are codons which are under-represented in a genome³². There are 15 preferred codons (RSCU value > 1.0) and 11 over-represented codons (RSCU value > 1.6) in SARS-CoV-2 genome as shown in **Table 2**. The preferred codons in SARS-CoV-2 genome are GCA (Ala), CGU (Arg), AAU (Asn), GAU (Asp), UGU (Cys), CAA (Gln), GAA (Glu), CAU (His), AUU (Ile), UUG (Leu), AAA (Lys), UUU (Phe), CCA (Pro), AGU (Ser) and UAU (Tyr) while the over-represented codons are GCU (Ala), AGA (Arg), GGU (Gly), CUU (Leu), UUA (Leu), CCU (Pro), UCA (Ser), UCU (Ser), ACA (Thr), ACU (Thr), and GUU (Val). The presence of the preferred and over-presented codons in a genome increases the protein synthesis rate.

Table 3 shows the RSCU analysis of the top 10 synonymous mutations. The codons in bold in the 'codon change' column are the codons with higher RSCU value, which means they are more preferred in SARS-CoV-2 genome. Most of the mutations change the codon to a more preferred codon as shown in **Table 3**. Since it is presumed that preferred codons have a higher translation rate compared to nonpreferred codons³⁴, it is possible that most of the mutations may increase the translation efficiency of SARS-CoV-2, which may affect virus replication, transmission, and evolution.

Table 3. RSCU analysis of the top 10 synonymous mutations of SARS-CoV-2 genome.

ORF	Mutation	Codon Change
1a	nsp1	C313U CUC -> CUU
	nsp2	C913U UCC -> UCU
	nsp3	C3037U UUC -> UUU
		C5986U UUC -> UUU
	nsp4	C9286U AAC -> AAU
1b	nsp12	C14676U CCC -> CCU
		C15279U CAC -> CAU
		U16176C ACU -> ACC
	nsp14	C18877U CUA -> UUA
M	C26735U	UAC -> UAU

Conclusions

The effects of SARS-CoV-2 synonymous mutations in various aspects such as RNA folding and RNA stability of the virus were studied, even though they do not cause changes in amino acid residue of the protein. Most of the synonymous mutations identified in the SARS-CoV-2 genome are found to have a minor effect on RNA folding and RNA stability of the virus except for C913U and C26735U mutations. Due to the shortcomings of prediction tools, experimental studies are needed to give a more comprehensive understanding of the biological consequences of synonymous mutations on SARS-CoV-2 virus.

Ethics and dissemination

No ethical approval is required for data analysis in this study (EA2702021).

Data and software availability

Underlying data

SARS-CoV-2 virus genome sequence data were obtained from the **GISAID Database**. The multiple alignment data can be assessed through FigShare.

Figshare: MSA (SARS-CoV-2). <https://doi.org/10.6084/m9.figshare.16681900.v1>³⁵

This project contains the following underlying data.

- MSA_0.fasta (multiple sequence alignment of SARS-CoV-2 sequences obtained between 31-12-2019 and 31-05-2020.)
- MSA_1.fasta (multiple sequence alignment of SARS-CoV-2 sequences obtained between 01-06-2020 and 15-10-2020.)
- MSA_2.fasta (multiple sequence alignment of SARS-CoV-2 sequences obtained between 16-10-2020 and 31-01-2021.)
- MSA_3.fasta (multiple sequence alignment of SARS-CoV-2 sequences obtained between 01-02-2021 to 22-03-2021.)

Data are available under the terms of the **Creative Commons Attribution 4.0 International license** (CC-BY 4.0).

Software

The python code for the identification of SARS-CoV-2 genome mutations can be assessed through **GitHub**.

Author contribution

CHN contributes to the concept, design, supervision of the project. WXB and SBZ contribute to the design, methodology, and data collection. WXB contributed to the analysis, and interpretation of data.

All authors were involved in drafting and revising the manuscript and approved the final version.

References

- Huang C, Wang Y, Li X, et al.: **Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.** *Lancet.* 2020; **395**(10223): 497–506.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wu D, Wu T, Liu Q, et al.: **The SARS-CoV-2 outbreak: What we know.** *Int J Infect Dis.* 2020; **94**: 44–48.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sharma A, Tiwari S, Deb MK, et al.: **Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): a global pandemic and treatment strategies.** *Int J Antimicrob Agents.* 2020; **56**(2): 106054.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sanjuán R, Domingo-Calap P.: **Mechanisms of viral mutation.** *Cell Mol Life Sci.* Birkhauser Verlag AG, 2016; **73**(23): 4433–4448.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kim JS, Jang JH, Kim JM, et al.: **Genome-Wide Identification and Characterization of Point Mutations in the SARS-CoV-2 Genome.** *Osong Public Heal Res Perspect.* 2020; **11**(3): 101–111.
[PubMed Abstract](#) | [Free Full Text](#)
- Khailany RA, Safdar M, Ozaslan M.: **Genomic characterization of a novel SARS-CoV-2.** *Gene Rep.* 2020; **19**: 100682.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bin Y, Wang X, Zhao L, et al.: **An analysis of mutational signatures of synonymous mutations across 15 cancer types.** *BMC Med Genet.* 2019; **20**(Suppl 2): 190.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sharp PM, Averof M, Lloyd AT, et al.: **DNA sequence evolution: the sounds of silence.** *Philos Trans R Soc Lond B Biol Sci.* 1995; **349**(1329): 241–247.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chamary JV, Parmley JL, Hurst LD.: **Hearing silence: Non-neutral evolution at synonymous sites in mammals.** *Nat Rev Genet.* 2006; **7**(2): 98–108.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Burrill CP, Westesson O, Schulte MB, et al.: **Global RNA Structure Analysis of Poliovirus Identifies a Conserved RNA Structure Involved in Viral Replication and Infectivity.** *J Virol.* 2013; **87**(21): 11670–11683.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mueller S, Papamichail D, Coleman JR, et al.: **Reduction of the Rate of Poliovirus Protein Synthesis through Large-Scale Codon Deoptimization Causes Attenuation of Viral Virulence by Lowering Specific Infectivity.** *J Virol.* 2006; **80**(19): 9687–9696.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lauring AS, Acevedo A, Cooper SB, et al.: **Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus.** *Cell Host Microbe.* 2012; **12**(5): 623–632.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- GISAID Initiative. [Accessed: 23-Sep-2021].
[Reference Source](#)
- Wu F, Zhao S, Yu B, et al.: **A new coronavirus associated with human respiratory disease in China.** *Nature.* 2020; **579**(7798): 265–269.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- MAFFT - a multiple sequence alignment program. [Accessed: 23-Sep-2021].
[Reference Source](#)
- Kumar S, Stecher G, Li M, et al.: **MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.** *Mol Biol Evol.* 2018; **35**(6): 1547–1549.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Miladi M, Raden M, Diederichs S, et al.: **MutaRNA: analysis and visualization of mutation-induced changes in RNA structure.** *Nucleic Acids Res.* 2020; **48**(W1): W287–W291.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sharma Y, Miladi M, Dukare S, et al.: **A pan-cancer analysis of synonymous mutations.** *Nat Commun.* 2019; **10**(1): 2569.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mochizuki T, Ohara R, Roossinck MJ.: **Large-Scale Synonymous Substitutions in Cucumber Mosaic Virus RNA 3 Facilitate Amino Acid Mutations in the Coat Protein.** *J Virol.* 2018; **92**(22): e01007–18.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sauna ZE, Kimchi-Sarfaty C.: **Synonymous Mutations as a Cause of Human Genetic Disease.** In *eLS*. Chichester, UK: John Wiley & Sons, Ltd, 2013.
[Publisher Full Text](#)
- Rice AM, Castillo Morales A, Ho AT, et al.: **Evidence for Strong Mutation Bias toward, and Selection against, U Content in SARS-CoV-2: Implications for Vaccine Design.** *Mol Biol Evol.* 2021; **38**(1): 67–83.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Simmonds P.: **Rampant C→U Hypermutation in the Genomes of SARS-CoV-2 and Other Coronaviruses: Causes and Consequences for Their Short- and Long-Term Evolutionary Trajectories.** *msphere.* 2020; **5**(3): e00408–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- De Maio N, Walker CR, Turakhia Y, et al.: **Mutation Rates and Selection on Synonymous Mutations in SARS-CoV-2.** *Genome Biol Evol.* 2021; **13**(5): evab087.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vandelli A, Monti M, Milanetti E, et al.: **Structural analysis of SARS-CoV-2 genome and predictions of the human interactome.** *Nucleic Acids Res.* 2020; **48**(20): 11270–11283.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ziv O, Price J, Shalamova L, et al.: **The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2.** *Mol Cell.* 2020; **80**(6): 1067–1077.e5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bhatt PR, Scaiola A, Loughran G, et al.: **Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome.** *Science.* 2021; **372**(6548): 1306–1313.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kelly JA, Olson AN, Neupane K, et al.: **Structural and functional conservation of the programmed –1 ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2).** *J Biol Chem.* 2020; **295**(31): 10741–10748.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cao C, Cai Z, Xiao X, et al.: **The architecture of the SARS-CoV-2 RNA genome inside virion.** *Nat Commun.* 2021; **12**(1): 3917.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cornillez-Ty CT, Liao L, Yates JR 3rd, et al.: **Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts with a Host Protein Complex Involved in Mitochondrial Biogenesis and Intracellular Signaling.** *J Virol.* 2009; **83**(19): 10314–10318.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sui L, Zhao Y, Wang W, et al.: **SARS-CoV-2 Membrane Protein Inhibits Type I Interferon Production Through Ubiquitin-Mediated Degradation of TBK1.** *Front Immunol.* 2021; **12**: 662989.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Faure G, Ogurtsov AY, Shabalina SA, et al.: **Role of mRNA structure in the control of protein folding.** *Nucleic Acids Res.* 2016; **44**(22): 10898–10911.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wong EH, Smith DK, Rabadan R, et al.: **Codon usage bias and the evolution of influenza A viruses. Codon Usage Biases of Influenza Virus.** *BMC Evol Biol.* 2010; **10**(1): 253.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sharp PM, Tuohy TM, Mosurski KR.: **Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes.** *Nucleic Acids Res.* 1986; **14**(13): 5125–5143.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Komar AA.: **The Yin and Yang of codon usage.** *Hum Mol Genet.* Oxford University Press, 2016; **25**(R2): R77–R85.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wx B.: **MSA (SARS-CoV-2).** *figshare.* Dataset. 2021.
<http://www.doi.org/10.6084/m9.figshare.16681900.v1>

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 28 April 2022

<https://doi.org/10.5256/f1000research.76505.r135257>

© 2022 Nithin C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chandran Nithin

Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland

The manuscript titled "Prediction of the Effects of Synonymous Variants on SARS-CoV-2 Genome" analyzes 30,299 SARS-CoV-2 genomes retrieved from GISAID for the period between 31st December 2019 and 22nd March 2021. The study identifies 381 mutations in the genome, of which 150 are synonymous. The study discusses two mutations, C913U and C26735U.

1. The dataset used in this study is outdated and represents only a short period between 31st December 2019 and 22nd March 2021. It will be necessary to update the data to include more sequences from a wider time frame.

The authors had indicated (through the editorial manager) that they are unwilling to update the dataset; the authors need to indicate the same in the title and detail the limitations of the dataset in the introduction section.

2. The authors have used the RNAfold program to predict the secondary structure of wild-type and mutant sequences. The authors need to use additional prediction tools and check for consensus between the predicted structure. Moreover, authors need to employ prediction tools to predict pseudoknots in the RNA secondary structure.
3. The secondary structure of SARS-CoV2 determined with the help of data from SHAPE-MaP, icSHAPE, and DMS-MaPseq experiments is available in the literature.^{1,2,3} The authors should utilize this data to augment this study.
4. The study requires the inclusion of appropriate controls which quantify the effect of mutations on the stability of the RNA structure. The authors state that "The minimum free energy value of the mutant (- 146.90 kcal/mol) is slightly less negative than that of the wild type (- 147.40 kcal/mol), which makes it a less thermodynamically stable structure compared to the wild type." This claim needs to be explored in detail and supported with evidence from experimental data and literature.

5. The authors state that C26735U mutation induces changes in the predicted RNA secondary structure by forming an extra multibranch loop at the mutation site. The authors should compare the same with secondary systems mentioned in point 3 (See Ref 1, 2, and 3). The authors need to perform a similar analysis for C913U.
6. The study identifies 381 mutations in the genome, of which 150 are synonymous; however, it discusses only two. The authors need to provide the data on the remaining 148 synonymous mutations and briefly discuss some of them in the manuscript.
7. The authors need to expand and include relevant literature in the introduction section.
8. The manuscript needs to be edited for language errors.

References

1. Lan T, Allan M, Malsick L, Woo J, et al.: Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. *Nature Communications*. 2022; **13** (1). [Publisher Full Text](#)
2. Huston N, Wan H, Strine M, de Cesaris Araujo Tavares R, et al.: Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Molecular Cell*. 2021; **81** (3): 584-598.e5 [Publisher Full Text](#)
3. Manfredonia I, Nithin C, Ponce-Salvatierra A, Ghosh P, et al.: Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Research*. 2020; **48** (22): 12436-12452 [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: RNA structure prediction; Computational Structural Biology; RNA-protein complexes.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 22 Aug 2022

Chong Han Ng

The manuscript titled "Prediction of the Effects of Synonymous Variants on SARS-CoV-2 Genome" analyzes 30,299 SARS-CoV-2 genomes retrieved from GISAID for the period between 31st December 2019 and 22nd March 2021. The study identifies 381 mutations in the genome, of which 150 are synonymous. The study discusses two mutations, C913U and C26735U.

1. The dataset used in this study is outdated and represents only a short period between 31st December 2019 and 22nd March 2021. It will be necessary to update the data to include more sequences from a wider time frame.

The authors had indicated (through the editorial manager) that they are unwilling to update the dataset; the authors need to indicate the same in the title and detail the limitations of the dataset in the introduction section.

Response: The landscape of mutational profile of SARS-CoV-2 genome is very dynamic, changing rapidly. As of 14 August 2022, there are more than 12 million SARS-CoV-2 virus genomes sequences deposited in GISAID database. Although it is useful to get more recent dataset, it is out of the scope of our study to update the data. Both identification of the SARS-CoV-2 mutations and the prediction analysis of the biological consequences of these mutations are very time-consuming processes. Based on the timelines of emergence of different variants, the cut-off date of our data collection overlaps with the period when the frequency of the alpha variant peaked. All the synonymous mutations except C241T reported in the alpha variant have been identified in our study as well. Although the alpha variant has been replaced by other variants in human population, it is detected from wildlife animals based on a few recent genomic surveillance analysis. It is possible that these alpha variant from wildlife animal may jump back to human, hence it remains relevant to predict the biological consequences of these synonymous mutations. The title of the paper has been revised accordingly. The limitation of the dataset was included in the discussion.

2. The authors have used the RNAfold program to predict the secondary structure of wild-type and mutant sequences. The authors need to use additional prediction tools and check for consensus between the predicted structure. Moreover, authors need to employ prediction tools to predict pseudoknots in the RNA secondary structure.

Response: To improve the outcome of the RNA secondary structure analysis, 3 prediction tools, namely RNAfold, IPknot++ and MXfold2 are used in our study. Of these, IPknot++ predicted some synonymous mutations may affect the pseudoknot formation. The prediction results for all top 10 synonymous mutations using these 3

tools are summarized in the Table 2

3. The secondary structure of SARS-CoV2 determined with the help of data from SHAPE-MaP, icSHAPE, and DMS-MaPseq experiments is available in the literature.^{1,2,3} The authors should utilize this data to augment this study

Response: To improve the outcome of the prediction analysis, the RNA secondary structure of wild type and mutant sequences were predicted using RNAfold program with the incorporation of SHAPE reactivity data obtained from the study done by Manfredonia *et al.* (2020).

4. The study requires the inclusion of appropriate controls which quantify the effect of mutations on the stability of the RNA structure. The authors state that "The minimum free energy value of the mutant (- 146.90 kcal/mol) is slightly less negative than that of the wild type (- 147.40 kcal/mol), which makes it a less thermodynamically stable structure compared to the wild type." This claim needs to be explored in detail and supported with evidence from experimental data and literature.

Response: Instead of comparing minimum free energy between the wild type and mutant, we applied three RNA secondary structure prediction tools to predict the effect of the mutation on RNA secondary structure.

5. The authors state that C26735U mutation induces changes in the predicted RNA secondary structure by forming an extra multibranch loop at the mutation site. The authors should compare the same with secondary systems mentioned in point 3 (See Ref 1, 2, and 3). The authors need to perform a similar analysis for C913U.

Response: The results of the 4 mutations showing changes in all 3 prediction tools are included in Figure 2-5.

6. The study identifies 381 mutations in the genome, of which 150 are synonymous; however, it discusses only two. The authors need to provide the data on the remaining 148 synonymous mutations and briefly discuss some of them in the manuscript.

Response: The RNA secondary structures prediction results using 3 prediction tools for the top 10 synonymous mutations have been included in Extended data 1-3. There were 4 mutations showing changes in all 3 prediction tools and the results are discussed in the manuscript.

7. The authors need to expand and include relevant literature in the introduction section.

Response: The relevant literature in the introduction section has been updated.

8. The manuscript needs to be edited for language errors.

Response: The language errors have been edited accordingly.

Competing Interests: No competing interests were disclosed.

Reviewer Report 20 December 2021

<https://doi.org/10.5256/f1000research.76505.r101828>

© 2021 Wang L. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Leyi Wang

Veterinary Diagnostic Laboratory and Department of Veterinary Clinical Medicine, College of Veterinary Medicine, University of Illinois, Urbana, IL, USA

The manuscript entitled “Prediction of the Effects of Synonymous Variants on SARS-CoV-2 Genome” reports an analysis of the SARS-CoV-2 genome obtained in the GISAID database. The authors analyzed synonymous changes of the 30,229 SARS-CoV-2 sequences available online, and the effects of these changes on RNA secondary structures and base-pair probability. They identified 150 synonymous mutations in 11 coding regions. Mutations with higher frequency are mostly found in ORF1a and ORF1b and two mutations (C913U and C26735U) found with effect on the predicted RNA secondary structure are not those with higher frequency.

This study collected 30,229 sequences in the database in the period of 31 Dec 2019 to 22 March 2021. As of 17 Dec 2021, there are over 4.4 million complete genomes with high coverage. Will these sequences used in the study include all variants reported or what sequences of variants are included? A table to summarize all variants with these mutations or not is needed.

A major concern is that the sequences even filtered by setting parameters to keep only sequencing with complete genome and high coverage still contain bad sequences, how did the authors exclude those sequences?

In the abstract section, the results part only included those mutations with high frequency but did not contain results of other analyses. Instead, they included the results of RNA secondary structure analysis in the conclusion.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Clinical Virology Diagnosis

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 22 Aug 2022

Chong Han Ng

The manuscript entitled "Prediction of the Effects of Synonymous Variants on SARS-CoV-2 Genome" reports an analysis of the SARS-CoV-2 genome obtained in the GISAID database. The authors analyzed synonymous changes of the 30,229 SARS-CoV-2 sequences available online, and the effects of these changes on RNA secondary structures and base-pair probability. They identified 150 synonymous mutations in 11 coding regions. Mutations with higher frequency are mostly found in ORF1a and ORF1b and two mutations (C913U and C26735U) found with effect on the predicted RNA secondary structure are not those with higher frequency.

This study collected 30,229 sequences in the database in the period of 31 Dec 2019 to 22 March 2021. As of 17 Dec 2021, there are over 4.4 million complete genomes with high coverage. Will these sequences used in the study include all variants reported or what sequences of variants are included? A table to summarize all variants with these mutations or not is needed.

Response: We downloaded the raw data on 23rd March 2021 and completed the data analysis in late June 2021. We submitted the manuscript in late August 2021. The landscape of mutational profile of SARS-CoV-2 genome is very dynamic, changing rapidly. It is far beyond the scope of this study to include all variants reported. Based on the timelines of emergence of different variants, the cut-off date of our data collection overlaps with the period when the frequency of the alpha variant peaked. All the synonymous mutations except C241T reported in the alpha strain have been identified in our study as well. We added new paragraph in the discussion section to discuss the impact of our findings and to explain why the study of alpha variant remains relevant now. In addition, we used 3 RNA

secondary prediction analysis tool, instead of one, to improve the outcome of the analysis. The prediction results for all top 10 synonymous mutations using these 3 tools and the base pair probability estimation results are summarized in the Table 2.

A major concern is that the sequences even filtered by setting parameters to keep only sequencing with complete genome and high coverage still contain bad sequences, how did the authors exclude those sequences?

Response: In our original method, we used the high coverage filter when downloading from GISAID database, in which only entries with less 1% N and 0.05% unique amino acids mutations are included. To further reduce bad sequences, we filtered to remove those sequences with higher than 0.1% N unresolved nucleotides and ambiguous letters. A total of 3584 sequences were removed by applying this filter. The list of the synonymous mutations remains unchanged despite of the revision of the mutation frequency.

In the abstract section, the results part only included those mutations with high frequency but did not contain results of other analyses. Instead, they included the results of RNA secondary structure analysis in the conclusion.

Response: The abstract has been revised to update the results and conclusion.

Competing Interests: No competing interests were disclosed.

Reviewer Report 02 November 2021

<https://doi.org/10.5256/f1000research.76505.r97299>

© 2021 Koyama T. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Takahiko Koyama

IBM TJ Watson Research Center, Yorktown Heights, NY, USA

Boon *et al.* made an investigation of synonymous mutations of SARS-CoV-2 genomes. Authors first identified common synonymous mutations such as 3037C>T followed by secondary structure predictions using RNAfold.

First, the number of genomes authors used is too small and too old. Currently, nearly 5 million genomes are uploaded at GISAID; however, authors used only 30,229. The set does not represent the current mutational landscape of SARS-CoV-2. Authors need to update the data. I have 3 million genomes analyzed and 18108 distinct synonymous mutations identified. The list of highest frequency in Table 1 looks completely different with the latest dataset.

Secondly, authors employed RNAfold for RNA secondary structure prediction. Although RNAfold is

a standard software for the task, the accuracy of the prediction is not impressive. Authors need to explore various other tools for the secondary structure or justify the usage of RNAfold.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome Analysis, SARS-CoV-2, Cancer, Immunology, Stem cell

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 04 Nov 2021

Chong Han Ng

Boon et al. made an investigation of synonymous mutations of SARS-CoV-2 genomes. Authors first identified common synonymous mutations such as 3037C>T followed by secondary structure predictions using RNAfold.

First, the number of genomes authors used is too small and too old. Currently, nearly 5 million genomes are uploaded at GISAID; however, authors used only 30,229. The set does not represent the current mutational landscape of SARS-CoV-2. Authors need to update the data. I have 3 million genomes analyzed and 18108 distinct synonymous mutations identified. The list of highest frequency in Table 1 looks completely different with the latest dataset.

RE: Thank you for the comments and feedbacks. We downloaded SARS-CoV-2 virus genome data submitted from 31 December 19 to 22 March 21 on 23 March 21. There were about a total of 842,603 SARS-CoV-2 virus genomes sequences before applying filters. SARS-CoV-2

genomic sequences were filtered by setting parameters to keep only sequences with complete genome, high coverage, low coverage excluded and patient status options. The use of the low coverage excluded and patient status filters were unintentionally left out in the manuscript and they will be added later. In our study, the use of these filters allows us to analyze high quality data to identify SARS-CoV-2 mutations without the requirement of high computing power. Our data is representative of the mutational landscape of SARS-CoV-2 for this specified period. Same synonymous mutations, including C913T, C3037T, C5986T, C14676T, C15279T and T16176C were identified in alpha strain by other research groups (<https://covariants.org/variants/20I.Alpha.V1>) and (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959438). The frequency of alpha strain peaked in April-May 21 and subsequently it has been quickly replaced by delta strain. Interestingly the defining synonymous mutations of delta strain (<https://covariants.org/variants/21A.Delta>) are very different from those of alpha strain. The mutational landscape of SARS-CoV-2 genome is very dynamic as shown by other research groups. Bear in mind that the main objective of our paper is not to monitor the mutational changes of SARS-CoV-2 genome. Although it is useful to get more recent dataset, it is out of the scope of our study to update the data. Both identification of the SARS-CoV-2 mutations and the prediction analysis of the biological consequences of these mutations are very time-consuming processes. The first draft of the manuscript was prepared in late June 21 and submitted in late August 21. Since different SARS-CoV-2 variants may have different sets of mutations, we will gain more insight to study the effects of different mutations in different SARS-CoV-2 variants. In our follow-up study, we plan to predict and compare the effect of the mutations of different variants. These analyses may provide some plausible explanation why some variants are replaced by other variants eventually.

Secondly, authors employed RNAfold for RNA secondary structure prediction. Although RNAfold is a standard software for the task, the accuracy of the prediction is not impressive. Authors need to explore various other tools for the secondary structure or justify the usage of RNAfold.

RE: We chose RNAfold because it is the most commonly used tools to predict the RNA secondary structure. Other than RNAfold, we also considered using Mfold to perform the RNA secondary structure prediction as well. However, the Mfold server was down during the time of the study. RNAfold computes an optimal structure over the whole sequence length, which means its performance and accuracy can be affected as the sequences used get longer. However, in our case, the accuracy of the prediction is still acceptable since the sequence we used to do the structure prediction is relatively short. Besides, the prediction results we obtained from RNAfold is further supported with the prediction results obtained from MutaRNA, which predicts the structural changes induced by the mutation by estimating the base pairing probabilities. In our results, the circular plot from MutaRNA shows the changes in the base pairing probabilities near the mutation site correlates well with the RNA secondary structure predicted by RNAfold.

Competing Interests: No competing interests were disclosed.

Author Response 21 Nov 2022

Chong Han Ng

First, the number of genomes authors used is too small and too old. Currently, nearly 5 million genomes are uploaded at GISAID; however, authors used only 30,229. The set does not represent the current mutational landscape of SARS-CoV-2. Authors need to update the data. I have 3 million genomes analyzed and 18108 distinct synonymous mutations identified. The list of highest frequency in Table 1 looks completely different with the latest dataset.

RE: We submitted a revised version in Sep 2022. Although we didn't use the latest dataset, we provide some justification in the discussion to explain why our data may remain relevant. We hope you will consider to review our manuscript again. The landscape of mutational profile of SARS-CoV-2 genome is very dynamic, changing rapidly. It is far beyond the scope of this study to include all variants reported. Based on the timelines of emergence of different variants, the cut-off date of our data collection overlaps with the period when the frequency of the alpha variant peaked. All the synonymous mutations except C241T reported in the alpha strain have been identified in our study as well. Although the alpha variant has been replaced by other variants in human population, it is detected from wildlife animals based on a few recent genomic surveillance analysis. It is possible that these alpha variant from wildlife animal may jump back to human, hence it remains relevant to predict the biological consequences of these synonymous mutations. The title of the paper has been revised accordingly. The limitation of the dataset was included in the discussion.

Secondly, authors employed RNAfold for RNA secondary structure prediction. Although RNAfold is a standard software for the task, the accuracy of the prediction is not impressive. Authors need to explore various other tools for the secondary structure or justify the usage of RNAfold.

RE: In the revised version, we used 3 RNA secondary prediction analysis tool, namely RNAfold, IPknot++, to improve the outcome of the analysis. The prediction results for all top 10 synonymous mutations using these 3 tools and the base pair probability estimation results are summarized in the Table 2.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research