Biomacromol. J., Vol. 7, No. 3, 174-184, December 2021.

Biomacromolecular Journal
www.bmmj.org

Comparative Analysis of Codon Usage, Rare Codon Clusters and Phylogenetic Relations of COVID-19, SARS, and MERS

M. Mortazavi[a], A. Malekpour[b], F. Kargar[c], S. Gholamzadeh[d,*], R. Mohammadi[e] and Z. Mohammadi[f]

[a]*Department of Biotechnology, Institute of Science and High Technology and Environmental Science, Graduate University of Advanced Technology, Kerman, Iran*

[b]*Legal Medicine Research Center, Legal Medicine Organization, Tehran, Iran*

[c]*Department of Medical Biotechnology, School of Advanced Medical Sciences and Technologies Tabriz University of Medical Sciences, Tabriz, Iran*

[d]*Saeid Gholamzadeh, Legal Medicine Research Center, Legal Medicine Organization, Vice Chancellor in Administration and Resources Development Affairs, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[e]*Genetic Laboratory of Shiraz Fertility Center, Shiraz, Iran*

[f]*Pathobiology laboratory of Ordibehesht Hospital, Shiraz, Iran*

(Received 9 October 2022, Accepted 24 February 2023)

ABSTRACT

Coronaviruses are enveloped single-stranded RNA genome viruses causing respiratory distress syndromes. The aim of this study was to compare codon usage, rare codon clusters, and phylogenetic relations in the orf1a polyprotein of COVID-19, SARS, and MERS coronaviruses. The frequency, number, and fraction of 61 codons for each amino acid were evaluated in the structure of viral protein and the preferred codons were assessed using the Gene Infinity website. The variations in codon usage bias were quantified by ENC and CBI in the ACUA software. Finally, the evolutionary relationship and phylogenetic analysis of Coronaviridae were studied using the MEGA 7 software. The GC3% of the cds was in the range of 15.668 to 16.534 and GC3 Skewness was from 0.299 to 0.34. The analysis of codon usage for all amino acids in SARS, MERS, and COVID-19 showed considerable differences between the three viruses. The close proximity of COVID-19 and SARS in the tree diagram represented a similarity in their gene sequence of orf1a polyprotein (pp1a). This phylogenetic analysis also indicated that COVID-19 and SARS had a near phylogenetic relation based on the nucleotide sequences of orf1a polyprotein in comparison to MERS. The findings of the present study revealed that the patterns of base compositions in COVID-19 are most likely the result of mutation pressure rather than that of natural selection since at all codon positions its effects are present. In addition, through the analysis of the base composition, it was found that the cds of COVID-19 are rich in AT, which should be considered in designing new drugs.

Keywords: Computational Biology, SARS, MERS and COVID-19

INTRODUCTION

Coronaviruses are pleomorphic RNA viruses that encompass crown-shaped peplomers belonging to the genus Coronavirus in the Coronaviridae family [1,2]. Coronaviruses with single-stranded positive-sense RNA genomes (32 kbp in length) are enveloped viral agents with the largest viral genomes [3-5]. Most Coronaviruses are not pathogenic for humans except for the causes of the severe acute respiratory syndrome (SARS) [6,7] and Middle East Respiratory Syndrome Coronavirus (MERS) [8-10]. Other most common types of coronaviruses, such as CoV OC43 and CoV 229E cause mild infections [11,12]. In fact, Coronaviruses as zoonotic viruses infect different animals and humans and cause numerous respiratory, gastrointestinal, hepatic, and neurologic clinical manifestations [4,13]. In December 2019, 27 viral pneumonia cases of pneumonia were found in Wuhan city originating from Huanan Seafood Market, and so, the first epidemic outbreak of COVID-19 was explored [14,15]. After a while, it was stated that

*Corresponding author. E-mail: Saeedghmail@yahoo.com

COVID-19 has been originated from wild bats and epigenetic assessments confirmed that such as SARS-CoV, the virus belongs to Group 2 of beta-coronavirus [16,17]. However, COVID-19 was found to have major differences in the outbreaks and clinical manifestations compared with SARS-CoV and MERS-CoV [18-20]. Whole genome analyses of the COVID-19 strain have provided major epigenetic findings and potential ways for discovering hidden information in the viral genome [21]. Computational study of COVID-19 may discover novel opportunities for developing novel effective drugs and vaccines. In addition, in-silico analysis of the genomes of COVID-19 and comparison with other major coronaviruses such as SARS-CoV and MERS-CoV will explore the secrets of similarities and differences in the pattern of the outbreak, morbidity, and mortality of mentioned Coronaviruses [15,17,22,23].

Today, an abundance of genomic data was provided from different organisms and recent studies have shown that synonymous codons are not used with the same frequency in organisms [24]. Characterization of the gene Coronaviridae family and possible their differences are likely to facilitate and contribute to the development of effective prevention and treatment protocols against coronavirus infection. The purpose of this bioinformatics study was to study the gene features of orf1a polyprotein (pp1a) in the Coronaviridae family (COVID-19, SARS, and MERS). In the present study, the CUB in the nucleotide sequence of orf1a polyproteins was analyzed by studying the codon bias index (CBI), codon adaptation index (CAI), and an effective number of codons (ENC). For further analysis, the three-dimensional structure of the nsp2 protein was modeled in the Robetta Server [25] and these structures were visualized and studied using Swiss PDB Viewer software [26] and PyMOL Molecular Graphics System [27]. More precise approaches can be chosen for treatment regimens using the findings of this study.

MATERIALS AND METHODS

Coronaviridae Genome Sequences

For bioinformatics' analysis, the nucleotide sequences and their features of the Coronaviridae family were obtained from NCBI (http://www.ncbi.nlm.nih.gov/) (Table 1).

Analysis of Codon Usage

Next, the frequency, number, and fraction of 61 codons for each amino acid were evaluated in the structure of Coronaviridae protein and the preferred codons were extracted using the Gene Infinity website (https://www.bioinformatics.org/sms2/codon_usage.html) [28] (Table 2).

Sequence Information and Annotation

The variation in codon usage bias was quantified by the ENC and CBI in the ACUA software [29]. To evaluate the codon usage bias, the ENC (value ranges from 20-61) is generally used [30]. Thus, the CBI (normalized between -1 and 1) also calculated the number of preferred codons that are used [31]. Thus, CBI value of zero means random choice is used, 1 means only preferred codons, and less than zero implies greater use of non-preferred codons [31].

In highly expressed genes, the CAI (ranges from 0 to 1) evaluates the degree of bias towards the codons [32]. On the gene in which uniformly all synonymous codons were used, the CAI would be 0 indicating no bias. In the other hand, in

Table 1. Genetic Properties of Coronaviridae (COVID-19, MERS, and SARS)

	COVID-19	MERS	SARS
Locus	29903 bp ss-RNA linear VRL 13-MAR-2020	30119 bp RNA linear VRL 13-AUG-2018	29751 bp ss-RNA linear VRL 13-AUG-2018
Accession	NC_045512	NC_019843	NC_004718
Version	NC_045512.2	NC_019843.3	NC_004718.3
Db_Xref	2697049	1335626	1489680
Protein ID:	YP_009725295.1	YP_009047203.1	NP_828850.1
Gene ID:	43740578	14254602	1489680

the genes where optimal codons were used, the CAI would be 1 for the strongest codon bias [33]. AT and GC content at three codon positions *i.e.*, AT1, AT2, AT3, GC1, GC2, and GC3 were calculated. GC3 content is assumed to be an excellent index of base composition bias [34].

Evolutionary Relationship and Protein Folding Rate

In the following, the evolutionary relationship and phylogenetic analysis of Coronaviridae were studied using the MEGA 7 software [29]. This analysis was performed by the construction of a phylogenetic tree with the maximum parsimony tree in MEGA 7. Frequencies of used codons were reported as descriptive statistics. For the protein folding rate study, the relative rareness of codons was studied in the %Min-Max algorithm [35].

Molecular Modeling of nsp2 Protein

To investigate the structural relationships, the 3D structure of nsp2 protein from COVID-19, SARS, and MERS was modeled by submission of nsp2 sequences in the Robetta Server [25]. The model which showed the best confidence and Z-score was selected and visualized using a Swiss PDB Viewer [36] and PyMOL molecular graphics system [27]. Hydrogen bonds were also calculated by PIC [37] and WHAT IF [38].

Compositional Analysis of Coronaviridae

The coding sequence of these viruses was retrieved and their codon usage bias and nucleotide composition bias were analyzed. The codon bias and nucleotide composition relationship were evaluated by comparing the values of A, T, G, C, and GC with the A3, T3, G3, C3, and GC3 values, respectively. The AT1, AT2, At3, GC1, GC2, and GC3 values were calculated for each gene to investigate the relationship between codon usage variation and compositional constraints. The GC3% of the cds was in the

Table 2. Compositional Analysis of Coronaviridae (MERS, COVID-19, and SARS) Genome Sequences. The Number of A, T, G, and C Nucleotides, Total bp, Percent of AT and GC, Codon Adaptation Index CAI, AT and GC Skewness, and Effective Number of Codons Enc

Gene	A	T	G	C	Total bp	AT%	GC%	AT Skewness	GC Skewness	CAI	ENc
MERS	3419	4263	2882	2612	13176	58.303	41.697	-0.11	0.049	0.664	50.155
COVID-19	3950	4281	2656	2331	13218	62.271	37.729	-0.04	0.065	0.647	44.513
SARS	3648	4061	2847	2593	13149	58.628	41.372	-0.054	0.047	0.674	48.236

Gene	A1	T1	G1	C1	AT1%	GC1%	AT1 Skewness	GC1 Skewness
MERS	1289	1367	685	1051	20.158	13.175	-0.029	-0.211
COVID-19	1381	1334	667	1024	20.54	12.793	0.017	-0.211
SARS	1351	1319	691	1022	20.306	13.028	0.012	-0.193

Gene	A2	T2	G2	C2	AT2%	GC2%	AT2 Skewness	GC2 Skewness
MERS	922	1867	753	850	21.167	12.166	-0.339	-0.061
COVID-19	1232	1949	595	630	24.066	9.268	-0.225	-0.029
SARS	1026	1804	744	809	21.523	11.811	-0.275	-0.042

Gene	A3	T3	G3	C3	AT3%	GC3%	AT3 Skewness	GC3 Skewness
MERS	1207	1029	1444	711	16.97	16.355	0.08	0.34
COVID-19	1336	998	1394	677	17.658	15.668	0.145	0.346
SARS	1270	938	1412	762	16.792	16.534	0.15	0.299

range of 15.668 to 16.534 and GC3 Skewness was from 0.299 to 0.34 (Table 2). The GC content at the first codon position (GC1) and second codon position (GC2) was compared with that of the third codon position (GC3) (Fig. 1). The results show that the patterns of base compositions are most likely the result of mutation pressure rather than that of natural selection since at all codon positions its effects are present [39].

Prevalence of Preferred (Used) Codons

Figure 1 shows the prevalence of preferred (used) codons in the Coronaviridae family. Here, it can be found which codon is preferred and used more than other codons. The results showed that some codon usage of amino acids was preferred in three different coronaviruses as the CCC for Pro. Also, some codon usage has the lowest preferred codon for these viruses as the GCG for Ala. However, some codons have different uses in these viruses as shown in these results.

The close proximity of COVID-19 and SARS in the tree diagram represented a similarity in their gene sequence of orf1a polyprotein (pp1a). This phylogenetic analysis also indicated that COVID-19 and SARS had a near phylogenetic distance based on the nucleotide sequences of the orf1a polyprotein (Fig. 1D).

To investigate the importance of the protein folding rate in the sequence of orf1a polyprotein, the relative rareness of codons was studied in the %Min-Max algorithm (Fig. 2) [35].

A sliding window of %Min-Max output along an mRNA sequence produces a plot in which clusters of predominantly common codons appear as positive (%Max) peaks, and clusters of predominantly rare codons appear as negative (%Min). The previous studies show that the synonymous rare codons can enhance co-translational protein folding, increasing the likelihood of forming the native protein structure and suppressing alternative folded structures [40, 41]. As these results show, COVID-19 has used a greater extent of the synonymous rare codons that can enhance folding yield (the fraction of proteins that fold and assemble to the native, active structure) (Fig. 2).

Molecular Modeling of nsp2 Protein

For the detailed study of structure, a 3D model of nsp2 protein was modeled in Robetta that created five models and the best model was selected, as shown in Fig. 3. Robetta is a protein structure prediction service that is continually evaluated through CAMEO [25]. The physicochemical properties of nsp2 were studied by the ProtParam tool (Table 3).

DISCUSSION

The genome of Coronaviruses' contains about ten open reading frames (ORFs) [42]. One of the essential ORFs is ORF1a/b which encodes two large polyproteins [43,44]. The pp1a and pp1ab polyproteins translated from ORF1a comprise the enzymes of the RNA-synthesizing complex involved in the viral genome replication and synthesis of subgenomic mRNA [45,46]. Thus, Orf1a plays a role in viral replication and transcription [47]. The present study was taken up to evaluate the bioinformatics parameters of codon usage bias along with the base composition of the coding sequences of Orf1a genes in the COVID-19, SARS, and MERS from the coronaviridae family.

Recently, computational biology approaches have been considered for the evaluation of different viral features and for discovering effective drugs [48-50]. In fact, bioinformatics studies explore hidden information and new insights about viral pathogenicity and provide the opportunity of finding potential effective treatment methods. In this study, this hidden information was evaluated and after a preliminary analysis of the base composition, it was found that the amount of AT is equal in the SARS and MERS but the percentage of AT is 4% more in COVID-19 in relation to the SARS and MERS. However, a complex correlation was observed while investigating the relationship between different nucleotides of the orf1a gene in the Coronaviridae family. The AT1 in the SARS, COVID-19, and MERS has an almost identical frequency. On the other hand, AT2 has a much higher frequency in COVID-19 in relation to SARS and MERS. AT3 has a slightly higher frequency in COVID-19. Moreover, the indices GC1, GC2, and GC3 were computed for each gene to establish the relationship among the three codon positions. GC1 has an almost identical frequency. On the other hand, the GC2 has a much lower frequency in COVID-19 and the AT3 has a slightly lower frequency in COVID-19. Our results showed that the coding sequences of the Coronaviridae family have a near range of GC3. Previous studies have shown that mutation pressure and

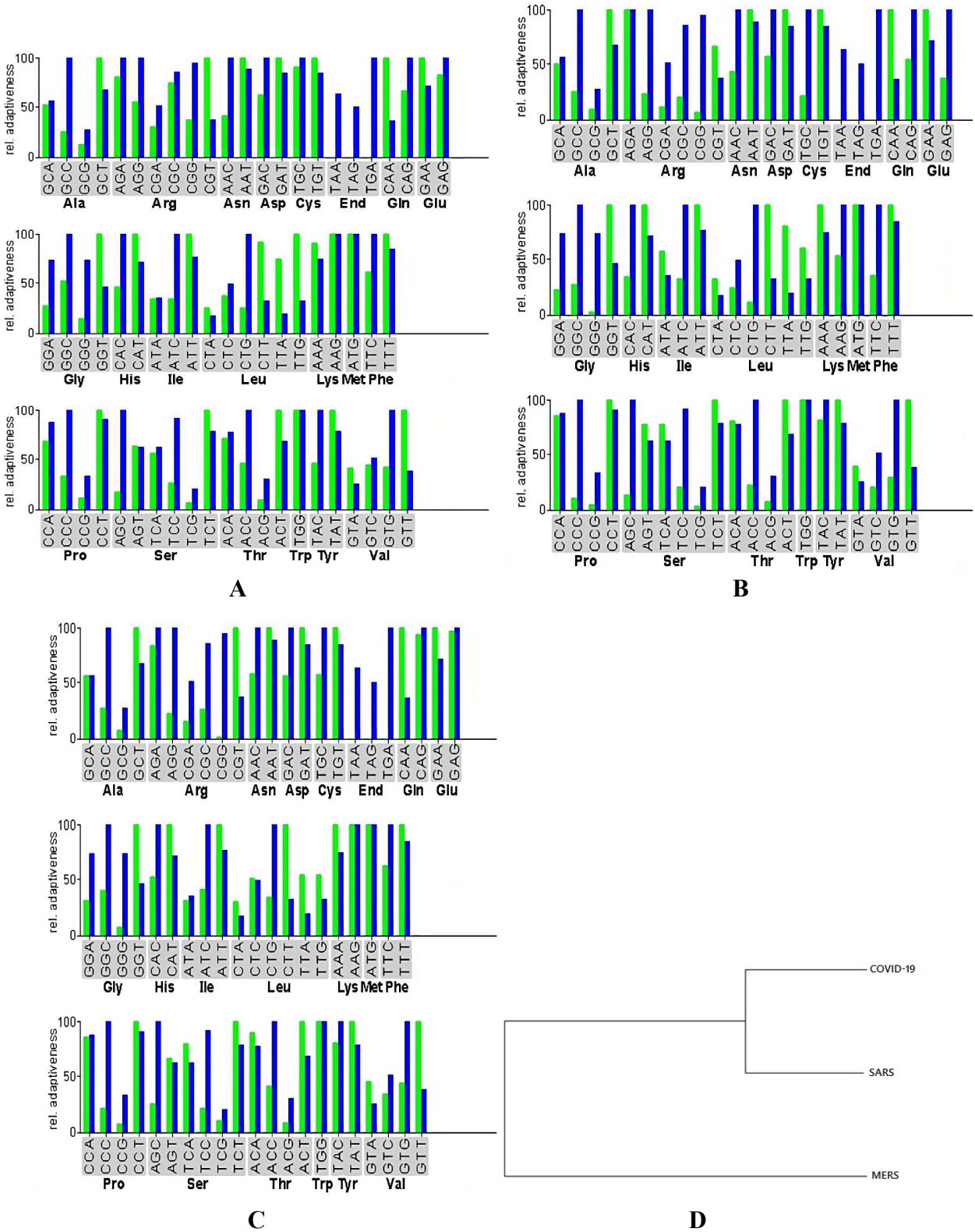
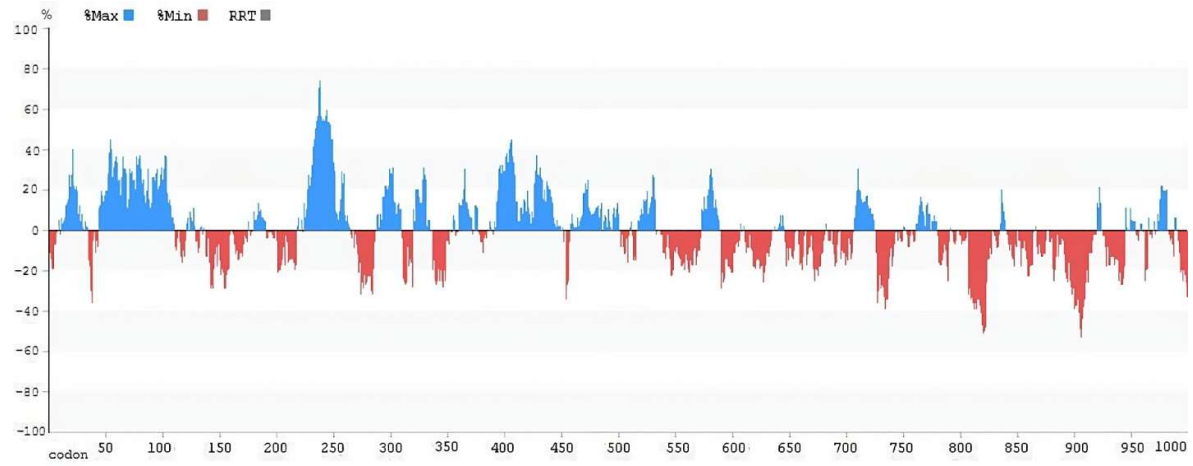
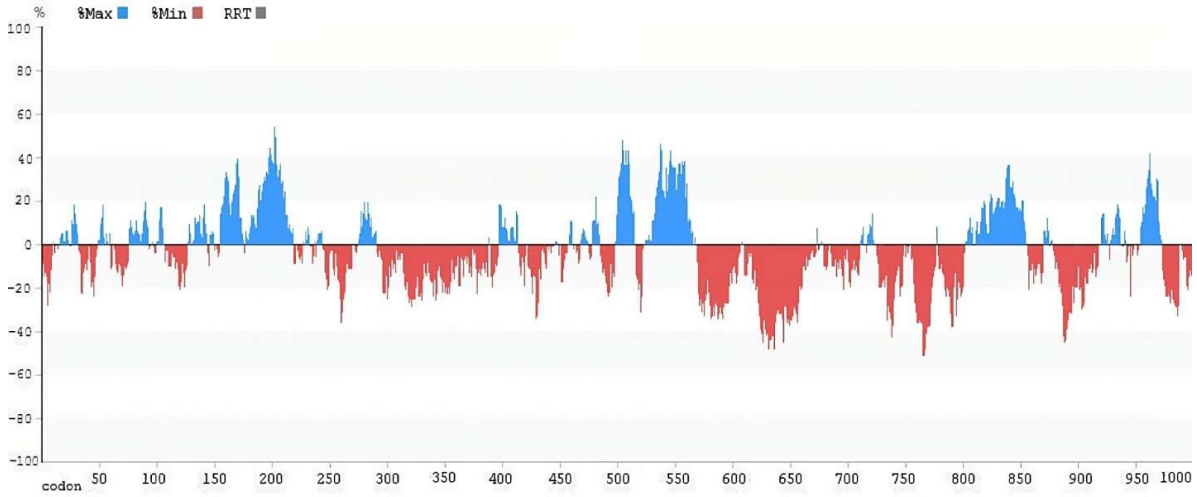


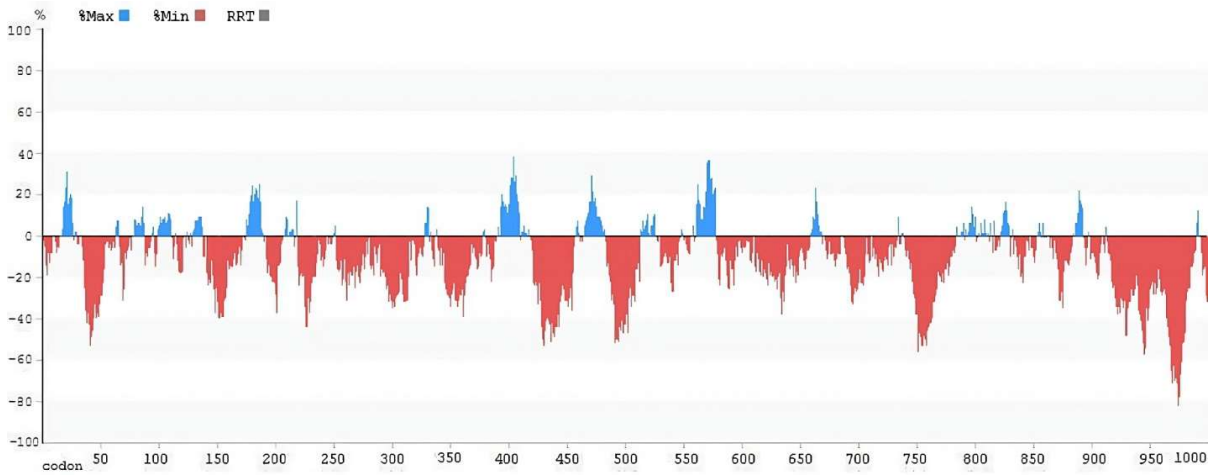
Fig. 1. The prevalence of preferred (used) codons in SARS (A), MERS (B), and COVID-19 viruses (C), D: Molecular Evolution and Phylogenetic Diagram Coronaviridae.



A

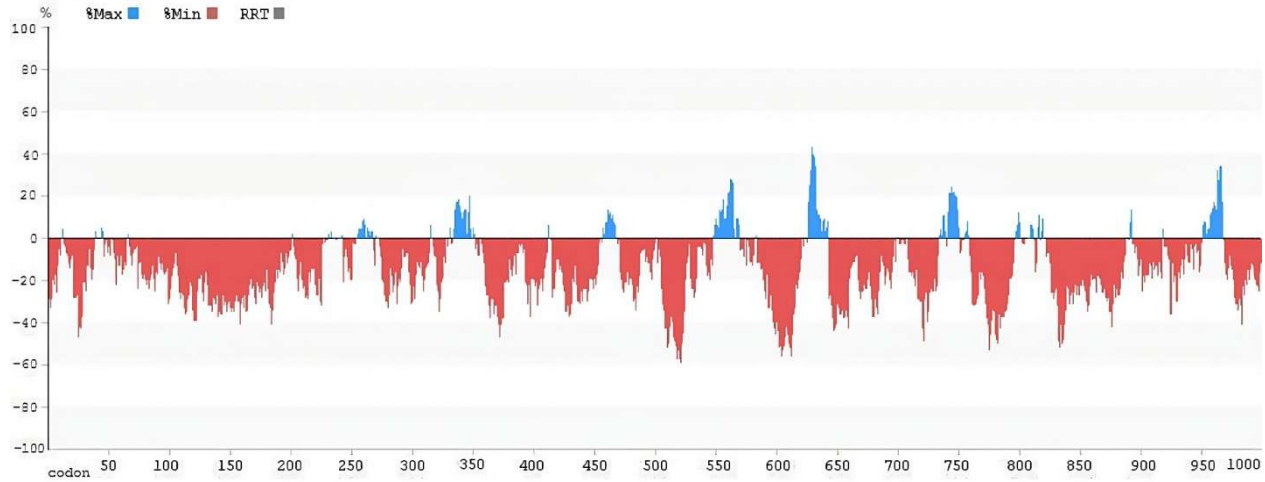


B

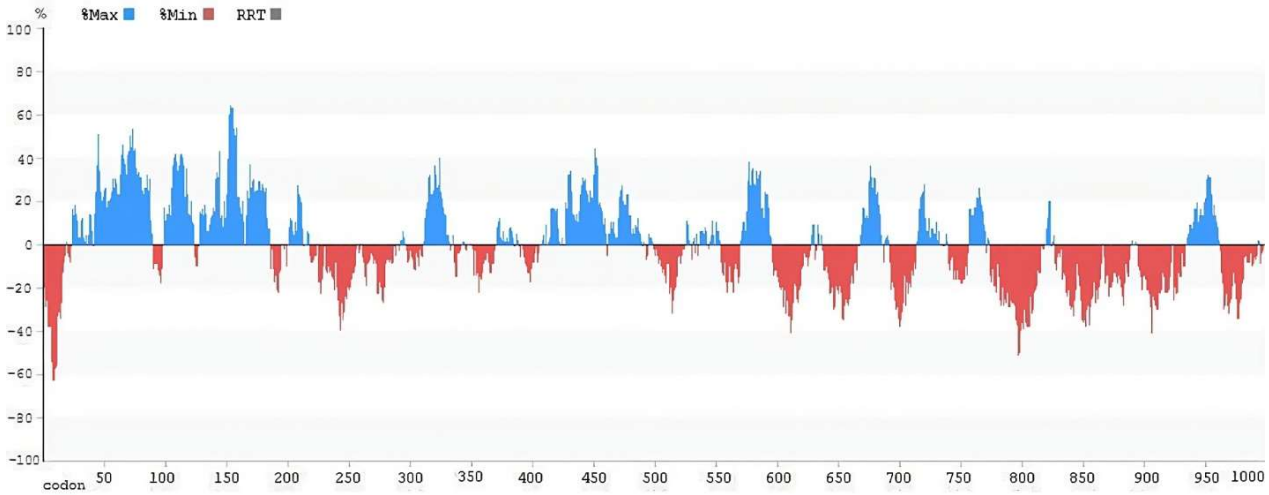


C

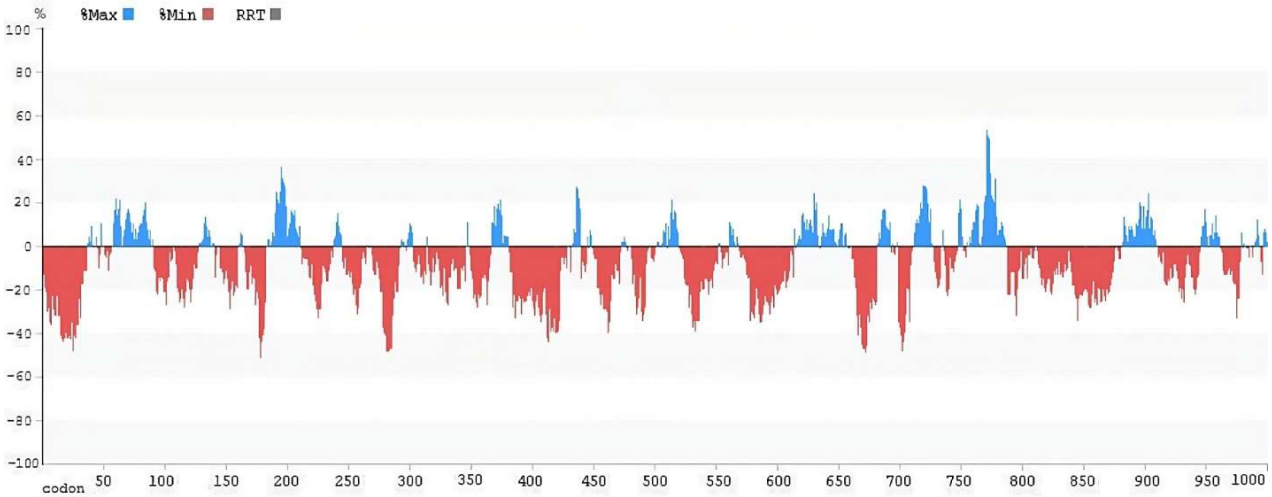
Fig. 2. The frequency and distribution of rare and common codons in the nucleotide sequence from 1 to 6000 in the SARS (A, B), COVID-19 (C, D), MERS (E, F).



D

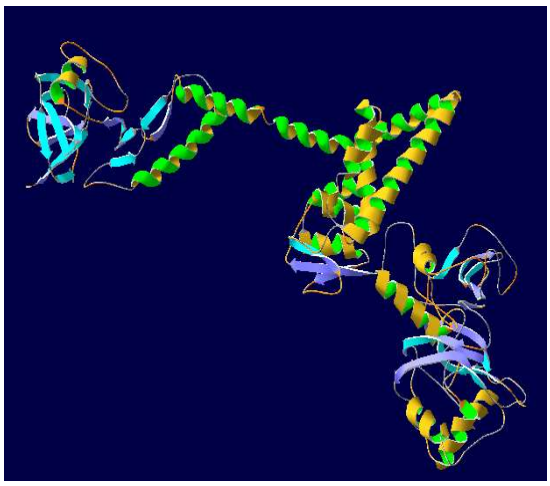


E

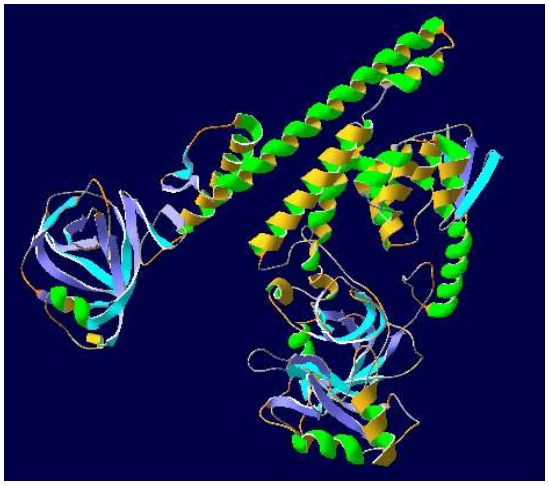


F

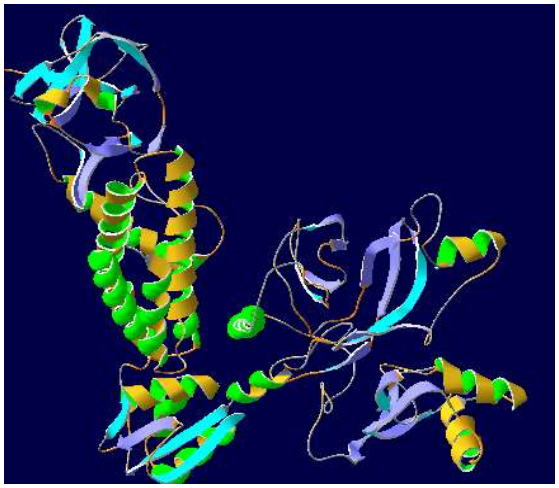
Fig. 2. Continued.



SARS



MERS



COVID-19

Table 3. *In-silico* Physicochemical Properties of the Nsp2 in the SARS, COVID-19, and MERS Obtained from ProtParam Tool. *First Value is Based on the Assumption that Both Cysteine Residues are Oxides and form Cystine and the Second Assumes that Both Cysteine Residues are Reduced

Parameters	SARS	MERS	COVID-19
Theoretical pI	6.08	5.7	6.25
Molecular weight	70622.10	72167.58	70511.38
Sequence length	638	660	638
Extinction coefficients (M^{-1} cm^{-1} at 260 nm)*	63830-65455	73690-74815	66810-68435
Asp + Glu	76	63	74
Arg + Lys	69	55	70
Instability index	35.96	31.24	36.06
Grand average of hydropathicity	89.51	0.114	-0.062
Aliphatic index	-0.109	91.21	88.93

natural selection are reasons for codon usage variation among different genes. If mutational pressure was the only effective reason for the codon usage bias, then the frequency of nucleotides C and G should be equal to A and T at the synonymous third codon position [51].

The orf1a gene located at the 5-terminus of the genome encodes the pp1a proteins. The pp1a protein encoded by the orf1a gene contains 10 nsps (nsp1-nsp10). Non-structural protein 2 (nsp2), is an RNA-binding protein that accumulates in cytoplasmic inclusions and is involved in coronavirus (CoVs) genome replication and suppressing gene expression. Compared to other nsp proteins, nsp2 varies among different CoVs strains. Different amino acid sequence identity has been observed in three groups (1, 2, and 3) of COVs. It is believed that nsp2 may mediate host-specific functions depending on the CoVs group. Because of its features, nsp2 protein is an attractive target for further genetic studies. On the other hand, to comprehend the structure of COVID-19, we modeled the structure of non-structural protein 2 from the COVID-19, SARS, and MERS using the I- the Robetta Server [25]. The results of modeling show that the structure of an nsp-2 protein is different in these COVID-19, SARS, and MERS. However, a slight similarity is observed between the nsp-2 protein from COVID-19 and MERS.

In our analysis, we have found that despite the high similarity of the nucleotide sequences of the orf1a polyprotein gene in the MERS, COVID-19, and SARS, the nucleotide sequences of the orf1a polyprotein gene in COVID-19 have some fundamental differences with the MERS and SARS. This analysis shows some variations in the nucleotide C and G with the A and T base compositions of COVID-19 in comparison to the MRES and SARS. This study reveals that other factors, such as natural selection or mysterious factors might also be the determining factor in shaping the synonymous codon usage pattern in COVID-19.

On the other hand, there is a reverse relationship between gene expression and ENC. The lower ENC value indicates a higher codon usage preference and higher gene expression [30]. Our results show that the ENC value in COVID-19 has lower that the MERS and SARS. This may indicate that COVID-19 has a higher gene expression in comparison with MRES and SARS. Similar to the ENC analysis, the low CAI value indicates a relatively high gene expression level. More interestingly, a correlation was observed between ENC and CAI (low for COVID-19) which shows that COVID-19 has a strong preference for a subset of codons (optimal codons and tRNA abundance) [52].

REFERENCES

[1] Fauci AS, Lane HC, Redfield RR. Covid-19—Navigating the Uncharted. In: Mass Medical Soc; 2020.

[2] Lipsitch M, Swerdlow DL, Finelli L. Defining the Epidemiology of Covid-19—Studies Needed. New England Journal of Medicine 2020.

[3] Weiss SR, Navas-Martin S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. Microbiol. Mol. Biol. Rev. 2005;69:635-664.

[4] Su S, Wong G, Shi W, Liu J, Lai AC, Zhou J, Liu W, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends in microbiology 2016;24:490-502.

[5] Cui J, Li F, Shi Z-L. Origin and evolution of pathogenic coronaviruses. Nature reviews Microbiology 2019;17:181-192.

[6] Kwok KO, Tang A, Wei VW, Park WH, Yeoh EK, Riley S. Epidemic Models of Contact Tracing: Systematic Review of Transmission Studies of Severe Acute Respiratory Syndrome and Middle East Respiratory Syndrome. Computational and structural biotechnology journal 2019.

[7] McIntosh K. Severe acute respiratory syndrome (SARS). In; 2016.

[8] Breban R, Riou J, Fontanet A. Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. The Lancet 2013;382:694-699.

[9] Drosten C, Seilmaier M, Corman VM, Hartmann W, Scheible G, Sack S, Guggemos W, et al. Clinical features and virological analysis of a case of Middle East respiratory syndrome coronavirus infection. The Lancet infectious diseases 2013;13:745-751.

[10] Chan JF, Lau SK, To KK, Cheng VC, Woo PC, Yuen K-Y. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. Clinical microbiology reviews 2015;28:465-522.

[11] Yin Y, Wunderink RG. MERS, SARS and other coronaviruses as causes of pneumonia. Respirology 2018;23:130-137.

[12] Peiris J, Lai S, Poon L, Guan Y, Yam L, Lim W, Nicholls J, et al. Coronavirus as a possible cause of severe acute respiratory syndrome. The Lancet 2003;361:1319-1325.

[13] Corman VM, Muth D, Niemeyer D, Drosten C: Hosts and sources of endemic human coronaviruses. In: Advances in virus research. Volume 100: Elsevier, 2018; 163-188.

[14] Read JM, Bridgen JR, Cummings DA, Ho A, Jewell CP. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. medRxiv 2020.

[15] Chan JF-W, Yuan S, Kok K-H, To KK-W, Chu H, Yang J, Xing F, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The Lancet 2020;395:514-523.

[16] Chen Z, Zhang W, Lu Y, Guo C, Guo Z, Liao C, Zhang X, et al. From SARS-CoV to Wuhan 2019-nCoV Outbreak: Similarity of Early Epidemic and Prediction of Future Trends. CELL-HOST-MICROBE-D-20-00063 2020.

[17] Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet 2020.

[18] Heymann DL, Shindo N. COVID-19: what is next for public health? The Lancet 2020.

[19] Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. Journal of Travel Medicine 2020.

[20] Mahase E. Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. In: British Medical Journal Publishing Group; 2020.

[21] Koyama T, Platt D, Parida L. Variant analysis of COVID-19 genomes.

[22] Dong N, Yang X, Ye L, Chen K, Chan EW-C, Yang M, Chen S. Genomic and protein structure modelling analysis depicts the origin and infectivity of 2019-nCoV, a new coronavirus which caused a pneumonia outbreak in Wuhan, China. bioRxiv 2020.

[23] Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Hu Y, Song Z-G, et al. Complete genome characterisation of a novel coronavirus associated with severe human respiratory disease in Wuhan, China. bioRxiv 2020.

[24] Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. Molecular biology and evolution 1985;2:13-34.

[25] Kim DE, Chivian D, Baker DJNar. Protein structure prediction and analysis using the Robetta server. 2004;32:W526-W531.

[26] Kaplan W, Littlejohn TG. Swiss-PDB viewer (deep view). Briefings in Bioinformatics 2001;2:195-197.

[27] DeLano WL. The PyMOL molecular graphics system. 2002.

[28] Stothard P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 2000;28:1102-1104.

[29] Vetrivel U, Arunkumar V, Dorairaj S. ACUA: a software tool for automated codon usage analysis. Bioinformation 2007;2:62.

[30] Wright F. The 'effective number of codons' used in a gene. Gene 1990;87:23-29.

[31] Bennetzen JL, Hall BD. Codon selection in yeast. Journal of Biological Chemistry 1982;257:3026-3031.

[32] Naya H, Romero H, Carels N, Zavala A, Musto H. Translational selection shapes codon usage in the GC-rich genome of Chlamydomonas reinhardtii. FEBS letters 2001;501:127-130.

[33] Stenico M, Lloyd AT, Sharp PM. Codon usage in Caenorhabditis elegans: delineation of translational selection and mutational biases. Nucleic acids research 1994;22:2437-2446.

[34] Zhou T, Gu W, Ma J, Sun X, Lu Z. Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. Biosystems 2005;81:77-86.

[35] Rodriguez A, Wright G, Emrich S, Clark PL. % MinMax: a versatile tool for calculating and comparing synonymous codon usage and its impact on protein folding. Protein Science 2018;27:356-362.

[36] Guex N, Peitsch M. Swiss-PdbViewer: a fast and easy-to-use PDB viewer for Macintosh and PC. Protein Data Bank Quaterly Newsletter 1996;77.

[37] Tina K, Bhadra R, Srinivasan N. Nucleic Acids Res 35.

Web Server issue) W473–W476 2007.

[38] Vriend G. WHAT IF: a molecular modeling and drug design program. Journal of molecular graphics 1990;8:52-56.

[39] SUPRIYO C, PROSENJIT P, MAZUMDER TH. Codon usage bias prefers AT bases in coding sequences among the essential genes of Haemophilus influenzae. Notulae Scientia Biologicae 2014;6:417-421.

[40] Lara HH, Ayala-Nuñez NV, Ixtepan-Turrent L, Rodriguez-Padilla C. Mode of antiviral action of silver nanoparticles against HIV-1. Journal of nanobiotechnology 2010;8:1.

[41] Jacobson GN, Clark PL. Quality over quantity: optimizing co-translational protein folding with non-'optimal'synonymous codons. Current opinion in structural biology 2016;38:102-110.

[42] Chan JF-W, Kok K-H, Zhu Z, Chu H, To KK-W, Yuan S, Yuen K-Y. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerging Microbes & Infections 2020;9:221-236.

[43] Li X, Geng M, Peng Y, Meng L, Lu S. Molecular immune pathogenesis and diagnosis of COVID-19. Journal of Pharmaceutical Analysis 2020.

[44] Enjuanes L. Coronavirus Replication and Reverse Genetics (Current Topics in Microbiology and Immunology). 2004.

[45] Snijder E, Decroly E, Ziebuhr J: The nonstructural proteins directing coronavirus RNA synthesis and processing. In: Advances in virus research. Volume 96: Elsevier, 2016; 59-126.

[46] Oudshoorn D, Rijs K, Limpens RW, Groen K, Koster AJ, Snijder EJ, Kikkert M, et al. Expression and cleavage of Middle East respiratory syndrome coronavirus nsp3-4 polyprotein induce the formation of double-membrane vesicles that mimic those associated with coronaviral RNA replication. MBio 2017;8:e01658-01617.

[47] Patino-Galindo JA, Filip I, AlQuraishi M, Rabadan R. Recombination and convergent evolution led to the emergence of 2019 Wuhan coronavirus. bioRxiv 2020.

[48] Brown JR, Magid-Slav M, Sanseau P, Rajpal DK. Computational biology approaches for selecting host–pathogen drug targets. Drug discovery today 2011;16:229-236.

[49] Fattahi M, Malekpour A, Mortazavi M, Safarpour A, Naseri N. The characteristics of rare codon clusters in the genome and proteins of hepatitis C virus; a bioinformatics look. Middle East journal of digestive diseases 2014;6:214.

[50] Mortazavi M, Zarenezhad M, Gholamzadeh S, Alavian SM, Ghorbani M, Dehghani R, Malekpour A, et al. Bioinformatic identification of rare codon clusters (RCCs) in HBV genome and evaluation of RCCs in proteins structure of Hepatitis B Virus. Hepatitis monthly 2016;16.

[51] Zhang Z, Dai W, Wang Y, Lu C, Fan H. Analysis of synonymous codon usage patterns in torque teno sus virus 1 (TTSuV1). Archives of virology 2013;158:145-154.

[52] Ikemura T. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. Journal of molecular biology 1981;151:389-409.