



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Research Paper

## Comparative analysis of protein synthesis rate in COVID-19 with other human coronaviruses

Chandra Mohan Dasari\*, Raju Bhukya (Ph.D) (Assistant Professor)

Department of Computer Science and Engineering, National Institute of Technology, Warangal, Telangana-506004, India

## ARTICLE INFO

**Keywords:**  
 Coronavirus  
 2019-nCoV  
 RSCU  
 COVID-19  
 Codon bias  
 Slow codons  
 Amino acid  
 ENC

## ABSTRACT

The genetic code contains information that impacts the efficiency and rate of translation. Translation elongation plays a crucial role in determining the composition of the proteome, errors within a protein contributes towards disease processes. It is important to analyze the novel coronavirus (2019-nCoV) at the codon level to find similarities and variations in hosts to compare with other human coronavirus (CoVs). This requires a comparative and comprehensive study of various human and zoonotic nature CoVs relating to codon usage bias, relative synonymous codon usage (RSCU), proportions of slow codons, and slow di-codons, the effective number of codons (ENC), mutation bias, codon adaptation index (CAI), and codon frequencies. In this work, seven different CoVs were analyzed to determine the protein synthesis rate and the adaptation of these viruses to the host cell. The result reveals that the proportions of slow codons and slow di-codons in human host of 2019-nCoV and SARS-CoV found to be similar and very less compared to the other five coronavirus types, which suggest that the 2019-nCoV and SARS-CoV have faster protein synthesis rate. Zoonotic CoVs have high RSCU and codon adaptation index than human CoVs which implies the high translation rate in zoonotic viruses. All CoVs have more AT% than GC% in genetic codon compositions. The average ENC values of seven CoVs ranged between 38.36 and 49.55, which implies the CoVs are highly conserved and are easily adapted to host cells. The mutation rate of 2019-nCoV is comparatively less than MERS-CoV and NL63 that shows an evidence for genetic diversity. Host-specific codon composition analysis portrays the relation between viral host sequences and the capability of novel virus replication in host cells. Moreover, the analysis provides useful measures for evaluating a virus-host adaptation, transmission potential of novel viruses, and thus contributes to the strategies of anti-viral drug design.

## 1. Introduction

The outbreak of Coronavirus 2019 (COVID-19) pandemic has led to a global emergency by affecting nearly 6.4 million people to date. The 2019-nCoV causes the severe acute respiratory syndrome, called as SARS-CoV-2, which infects the respiratory tract of humans (Association et al., 2020). There is no known vaccine and anti-viral treatment, primary treatment is based on symptoms and support therapy. CoVs are family of viruses classified into two groups as common human CoVs and other human CoVs. 229E, NL63, OC43, and HKU1 are common human CoVs that pose less threat to humans, whereas, other human CoVs like SARS-CoV, MERS-CoV, and SARS-CoV-2 are evolved from animals and were transmitted to humans which pose a greater threat to humans (Health et al., 2013). MERS-CoV, SARS-CoV, and SARS-CoV-2 are zoonotic in nature as they are originated from animals and transmitted to humans.

CoVs can infect only certain cells of the specific host of particular species. Host specificity is the number and identity of host species which are used by parasites (viruses). The molecular basis for host specificity indicates that a surface molecule called viral receptor needs to be located on surface of host for virus to bind. The host specificity of different CoVs not only limits to humans but also for various animals (bats, camels, pigs and dogs). The viruses use host receptors as an entry point to invade human cells and use their machinery to reproduce, such cells are called permissive (Zedalis and Eggebrecht, 2018). In humans, Angiotensin-Converting-Enzyme 2 (ACE2) acts as the host receptor for 2019-nCoV, SARS-CoV and NL63 (Hoffmann et al., 2020), Dipeptidyl Peptidase IV (DPP4) for MERS-CoV (Li, 2015), Aminopeptidase N (hAPN) for 229E (Li, 2015), N-acetyl-9-O-acetylneuraminic acid (Neu5,9Ac2) for OC43, and O-Acetylated Sialic acid for HKU1 (Krempl et al., 1995). CoVs are positive-stranded DNA (or) RNA genomes present inside a protein shell called Nucleocapsid (N) along with two

\* Corresponding author.

E-mail addresses: [chandu.dasari@student.nitw.ac.in](mailto:chandu.dasari@student.nitw.ac.in) (C.M. Dasari), [raju@nitw.ac.in](mailto:raju@nitw.ac.in) (R. Bhukya).<https://doi.org/10.1016/j.meegid.2020.104432>

Received 29 April 2020; Received in revised form 7 June 2020; Accepted 15 June 2020

Available online 25 June 2020

1567-1348/ © 2020 Elsevier B.V. All rights reserved.

membrane proteins, namely, Membrane (M), Envelope (E), and one glycoprotein Spike (S) (Siddell et al., 2010). The S protein attaches the virion to the receptor (ACE2) of the host cell by initiating the infection owing to drastic changes (Li, 2016). It is vulnerable to mutations on the receptor interface to defend the host immunity (Sui et al., 2014). The E protein forms protein-lipid pores and also acts as a channel for ion transport (Ruch and Machamer, 2012). The M protein plays a vital role in enhancing virus RNA transcription and helps in biosynthesis of the virus (Neuman et al., 2011). The N protein is the most stable and conserved multifunctional protein that interacts with M protein during virion assembly and controls the host cells by intoxicating cell machinery (McBride et al., 2014). DNA viruses use proteins and enzymes of host cell to develop new virus DNA, later that is transcribed to messenger RNA. The viral mRNA controls the host cell to synthesize its proteins and enzymes.

The genetic code defines the relationship between codons and amino acids sequence of the protein. During translation, mRNA is read and decode for proteins that contain a chain of amino acids or polypeptides. The translation process is carried out in three stages initiation, elongation, and termination. Ribosomes and transfer RNA (tRNA) are two molecules play a major role during the translation process. In the initiation stage, ribosomes sit around the mRNA strand and the first tRNA carrying the amino acid and the anti-codon interacts with a start codon (ATG) of the mRNA template to start the translation. During translation elongation, the amino acid chain grows longer as each time a new codon is encountered mRNA binds with matching anti-codon of its cognate tRNA on the A-site of the ribosome. The existing polypeptide chain links onto the amino acid of tRNA on P-site via a chemical reaction. Then mRNA is shifted one codon to make room for new codon to decode. Finally in the termination stage decoding stops when stop codons (TAG, TAA, or TGA) are encountered and finished polypeptide chain is released. Based on the behaviour of antisense molecules, new anti-viral drugs have been discovered (Stein et al., 2001). Antiviral drugs stall the translation process thereby stopping the protein synthesis of the virus and end its replication.

There are 64 codons (61 codons for decoding and remaining 3 used as stop codons to terminate translation) which code for 20 amino acids in protein synthesis. The translation process has become complicated because more than one codon decodes the same amino acid such codons are referred as synonymous codons. This state of genetic code is indicated as 'degeneracy' of the codons (Gonzalez et al., 2019). Various studies revealed that the selection of synonymous codons is a non-random procedure (Plotkin and Kudla, 2011; Sharp et al., 2010). In translation some codons are preferred over other synonymous codons is known to be 'codon usage bias' or 'codon bias'. Some unknown codons or combination of codons influence the translation rate and efficiency, such codons act as rate-limiting factors. Moreover, it is difficult to comprehend how much difference in translation rate, affects the gene expression (Brule and Grayhack, 2017). Depending upon the species usually, there are 40 to 60 different types of tRNAs with its anti-codon and matching amino acid exists in a cell. The codon usage pattern analysis in various CoVs may provide useful insights in the synonymous codon usage and host-adapted evolutionary process.

The ribosome decodes the 61 codons to 20 amino acids at non-uniform rate is referred as codon optimality (Hanson and Coller, 2018). It acts as a strong cognitive factor for translation rates, the corresponding genes seems to be have inherent codon bias. The influence of codon optimality on the translation rate remains a topic of potent debate (Novoa and de Poupplana, 2012). The concern debate, on codon optimality focused much around the influence of codon bias on translation efficiency, protein folding and translation fidelity. More recent studies reveals that the codon optimality became an important determinant of translation elongation rate and mRNA stability (Presnyak et al., 2015; Harigaya and Parker, 2016; Radhakrishnan et al., 2016). For over two decades, it was assumed that various codons were translated at different rate. It was evident that codon bias could affect the

translation elongation speed by the use of a radio-labelled amino acid incorporation assay (Sørensen and Pedersen, 1991). Sometimes ribosomes are forced to wait for longer duration due to rare cognate tRNA to bind on ribosome A-site. This situation leads to per-codon elongation rate as they are majorly dependent on tRNA pool (Koutmou et al., 2015). The overhead of kinetic cost levied to elongation rate for filtering and rejecting the near-cognate tRNA species where the cognate tRNA are scarce (Chu et al., 2011). The abundance of codon usage and tRNA increase the gene expression along with regulating ribosomal translation speed, protein folding efficiency and gene expression functionalities (Novoa and de Poupplana, 2012). The recent emerging methods by monitoring translation in real time provides additional evidence for the influence of codon bias on translation elongation rates (Chekulaeva and Landthaler, 2016; Iwasaki and Ingolia, 2016).

Codons without corresponding cognate tRNA are known as slow codons of the organism (Yang and Chen, 2020). The combination of two consecutive slow codons is called slow dicodons (Yang and Chen, 2020). The human genome contains 13 slow codons which have no cognate tRNA genes (Chan and Lowe, 2009). Ribosomal experiments conducted in many studies show that decoding codons with non-cognate tRNAs and rare tRNAs greatly reduce the translation efficiency (Dana and Tuller, 2012; Hussmann et al., 2015; Stadler and Fire, 2011). Furthermore, it was reported that the decoding efficiency of a particular codon is influenced by its subsequent codon. Two subsequent slow codons (slow di-codons) reduces translation rate (Chevance and Hughes, 2017; Chevance et al., 2014). Conversely, the low proportions of slow codons and slow di-codons may have rapid rate of translation in the mRNA coding sequences (CDs). Viruses use the human host cell translation machinery for their protein translation. Therefore, host-specific slow codons and slow di-codons proportions in the viral host sequences can be used to anticipate the order of translation rates between various viruses of different genera, serotypes, and strains.

Human CoVs were first recognized in children during 1960 are responsible for upper respiratory tract infections. In 1965 a virus B814 was discovered in adults and later renamed as 229E (Tyrrell et al., 1966). Later OC43, beta coronavirus was found which spreads over humans and cattle. SARS-CoV first recognized in Guangdong province, China in November 2002. This infectious disease said to report 8000 cases with the death of 800 people according to the report by the World Health Organization (WHO). Thereafter another coronavirus, NL63 was found in 2004 that affect more severe lower respiratory or mild upper respiratory systems mostly common in younger children (Abdul-Rasool and Fielding, 2010). In 2005 HKU1 was identified as genetically distinct from OC43 and other CoVs (Kahn and McIntosh, 2005). MERS-CoV was a zoonotic virus first identified in 2012. Due to MERS-CoV, nearly 35% of infected humans have deceased according to the report of WHO. In December 2019, a novel coronavirus (COVID-19) which is homologous with SARS-CoV has created havoc worldwide infecting 6.4 million people in more than 200 countries till 05-06-2020.

In this study, we analyzed comprehensively various characteristics of seven different CoVs which includes codon usage bias, the ENC, RSCU, proportions of slow codons and slow di-codons, mutation bias, CAI, codon frequencies, and AT1-AT3 and GC1-GC3 count. All these factors influence the translational rate and efficiency. We aim to interpret the codon patterns of these seven CoVs concerning the host relation. The major highlights of this study include:

- 1) The proportions of human slow codons and slow di-codons of 2019-nCoV and SARS-CoV found to be similar and very less compared to the other five coronavirus types, which suggest that the 2019-nCoV has faster protein synthesis rate.
- 2) The zoonotic nature CoVs 2019-nCoV, SARS-CoV, and MERS-CoV has fewer number of negatively biased (average RSCU value < 0.6) codons than four human CoVs, which implies the high translation rate in zoonotic viruses.
- 3) Most of the CoVs isolates and its clusters fall under the standard

curve, which strengthen the role of directional mutation pressure and natural selection on codon bias respectively.

- 4) It is observed that all the CoVs have more AT% than GC% in genetic codon compositions. The CAI values of SARS-CoV and MERS-CoV are maximum, so the rate of gene expression is high in these viruses.
- 5) In MERS-CoV and NL63, we observed enormous silent, missense and nonsense mutations whereas in 2019-nCoV there are a little silent mutations and none nonsense mutations.

## 2. Materials and methods

In this section, we collected datasets and introduced various parameters that influence codon usage bias which helps to compare and analyze different types of CoVs.

### 2.1. Data collection

The segment of the gene (DNA or RNA) that code for protein is a coding sequence. We have collected and analyzed 4143 CDs of seven types of coronavirus that can infect human hosts from the virus pathogen resource (ViPR) (Pickett et al., 2011). The collected CDs are host specific viral infected mRNAs and the details are given in the Table 1. The reference genomes and test sequences of various CoVs for evaluating the mutational bias are collected from NCBI. There are 13 human slow codons (ACC, AGT, CAT, CCC, CGC, CTC, GAT, GCC, GGT, GTC, TCC, TGT, TTT) (Yang and Chen, 2020). By the combinations of these 13 slow codons, a total of 169 slow di-codons were formed. For example, by combing ACC, AGT, and CAT we can form six slow di-codons ACCAGT, AGTACC, ACCCAT, CATACC, AGTCAT, and CATAGT. Two consecutive slow codons can reduce the translation rate extremely.

### 2.2. Data analysis parameters

We consider various parameters that are prominent for analyzing various coronavirus CDs.

#### 2.2.1. Relative synonymous codon usage

In most eukaryote (including human) and prokaryote species, for any given amino acid some codons are preferred over other synonymous codons in decoding. Many statistical methods exist to measure such codon bias. Relative synonymous codon usage is a widely used measure introduced (Sharp et al., 1986). For a given amino acid  $i$ , consider  $s_i$  be the number of synonymous codons that decode amino acid  $i$ . Let  $f_{i,j}$  denotes the frequency of synonymous codon  $j$  of  $i^{\text{th}}$  amino acid. Then the RSCU measure for codon  $j$  of  $i$  is calculated as follows:

$$RSCU_{i,j} = \frac{s_i * f_{i,j}}{\sum_{j=1}^{s_i} f_{i,j}} \quad (1)$$

RSCU is measured for each amino acid synonymous codon. It ranges between 0 and the number of synonymous codons of particular amino acids. High RSCU values of specific codons indicate that codon plays an important role in protein synthesis.

**Table 1**

Number of coding sequences, its ranges and discovered years of different coronaviruses.

Corona Type	Year	No. of CDs	Range(bps)
229E	1965	113	51–20,292
OC43	1980	1130	239–21,288
SARS-CoV	2002	1863	80–21,291
NL63	2004	270	51–20,190
HKU1	2005	238	249–21,654
MERS-CoV	2012	27	42–21,237
2019-nCoV	2019	502	107–21,291

#### 2.2.2. Effective number of codons

Another metric formulated to measure codon usage bias is an effective number of codons (Wright, 1990; Sun et al., 2013). Depending on the number of synonymous codons, the amino acids are classified into five synonymous codon family ( $S_{cf}$ ) types (1,2,3,4, 6).  $S_{cf}$ -type1 has 2 amino acids with one codon choice,  $S_{cf}$ -type2 has 9 with 2, type3 has 1 with 3, type4 has 5 with 4 and type6 has 3 with 6 codon choices. For individual  $S_{cf}$  of  $j$  codons whose actual usage are  $n_1, n_2, \dots, n_j$  ( $j \in \{1, 2, 3, 4, 6\}$ ), and the total usage is  $n = \sum n_i, i \in 1$  to  $m$ . The frequency usage of codons are  $p_1, p_2, \dots, p_j$  ( $p_i = n_i / n$ ). Homozygosity of particular  $S_{cf}$  can be calculated from the squared codon frequency

$$F_{S_{cf}} = \frac{n \sum_{i=1}^j p_i^2 - 1}{n - 1} \quad (n \neq 1) \quad (2)$$

Effective number of codons for particular  $S_{cf}$  type can be measured as follows:

$$N_{S_{cf}} = \frac{1}{F_{S_{cf}}} \quad (3)$$

The effective number of codons is measured by adding contributions from each of the  $S_{cf}$

$$ENC = 2 + 9 * \check{N}_{S_{cf2}} + 1 * \check{N}_{S_{cf3}} + 5 * \check{N}_{S_{cf4}} + 3 * \check{N}_{S_{cf6}} \quad (4)$$

where  $\check{N}_{S_{cfi}}$  is average homozygosity for synonymous codon family type  $i, i \in \{2,3,4,6\}$ . The lower ENC values of CoVs suggest higher gene expression efficiency (Kandee et al., 2020).

#### 2.2.3. Proportion of slow codons and slow di-codons

The proportion of human slow codons for a given coronavirus CDs is measured as follows:

$$P_{sc} = \sum_{i=1}^n \frac{N_{sc}}{T_c} \quad (5)$$

Where  $N_{sc}$  is the number of slow codons present in each virus coding sequence.  $T_c$  represents the total number of codons and  $n$  represents the total number of coronavirus CDs.

The proportion of human slow di-codons for a given coronavirus CDs is calculated as follows:

$$P_{sdc} = \sum_{i=1}^n \frac{N_{sdc}}{T_c - 1} \quad (6)$$

where  $N_{sdc}$  is the number of slow di-codons present in each virus coding sequence.  $T_c$  represents the total number of codons and  $n$  represents the total number of coronavirus CDs. A virus depends completely on the host cell machinery to replicate by undergoing a vital process of translation. Consequently the proportions of host-specific slow codons and slow di-codons can be considered to predict the order of the viral mRNA protein synthesis rates of various viruses of different genus and strains. Lower the proportions of slow codons and slow di-codons may have fast translation rates in the CoVs CDs.

#### 2.2.4. Mutation bias

Mutations can occur randomly but all these mutations are not equally likely. Mutation bias indicates a portion of sequence in which particular type of mutation occur more frequently but the exact cause is unknown (Yampolsky and Stoltzfus, 2001). Transition-Transversion bias is a type of mutational bias. Substitution mutations are classified as transitions and transversions. Based on the structure the nucleotides A, G and C, T are called purines (two-ring) and pyrimidines (one-ring) respectively. The transition mutation occurs due to mutations among purines ( $A \leftrightarrow G$ ) and pyrimidines ( $C \leftrightarrow T$ ) hence, the mutation involves between similar shape (Keller et al., 2007). Whereas, in transversions the mutations occur between purine (A, G) and pyrimidine (C, T) bases. Therefore in transversions the mutations involve exchange of bases between one-ring and two-ring structures. The number of possible choices in transversions is relatively higher than transitions even though the transitions are at higher frequency

**Table 2**  
RSCU values of different types of coronaviruses.

Amino Acid	Codon	229E	HKU1	NL63	OC43	MERS-CoV	SAR-COV	2019-nCoV
Ala	GCT	<b>2.178</b>	<b>2.398</b>	<b>2.308</b>	<b>2.218</b>	<b>2.032</b>	<b>2.011</b>	<b>2.022</b>
	GCC	0.717	0.694	0.586	0.602	0.705	0.718	0.652
	GCA	1.132	0.876	1.159	1.159	0.942	1.545	1.556
Arg	GCG	0.304	0.359	0.192	0.313	0.511	0.561	0.533
	CGA	0.818	0.553	0.763	0.761	0.935	0.922	0.915
	CGC	0.697	1.051	0.761	0.995	1.284	0.762	0.716
Asn	CGT	<b>1.973</b>	<b>2.308</b>	<b>2.695</b>	<b>2.209</b>	<b>2.092</b>	<b>1.797</b>	<b>1.807</b>
	CGG	0.561	0.517	0.463	0.528	0.856	0.334	0.3
	AGA	<b>2.621</b>	<b>2.192</b>	<b>1.623</b>	<b>2.079</b>	<b>1.626</b>	<b>2.739</b>	<b>2.742</b>
Asp	AGG	0.909	0.857	1.172	1.16	1.467	1.775	1.765
	AAC	0.699	0.339	0.513	0.497	0.655	0.809	0.807
	AAT	1.37	<b>1.691</b>	<b>1.639</b>	1.59	1.37	1.281	1.274
Cys	GAC	0.778	0.493	0.577	0.502	0.796	0.895	0.895
	GAT	1.244	<b>1.677</b>	1.486	1.551	1.281	1.28	1.289
	TGC	0.689	0.361	0.435	0.641	0.827	0.993	0.999
Gln	TGT	1.438	<b>1.737</b>	<b>1.768</b>	1.462	1.243	1.296	1.287
	CAA	1.288	1.32	1.324	1.167	1.135	1.394	1.384
	CAG	0.827	0.767	0.773	0.881	0.865	0.695	0.694
Glu	GAA	1.195	1.411	1.402	1.237	1.005	1.286	1.278
	GAG	0.871	0.716	0.869	0.879	1.107	0.733	0.723
	GGA	0.542	0.577	0.488	0.741	0.87	1.169	1.176
Gly	GGC	1.012	0.709	0.652	0.701	1.04	1.482	1.491
	GGT	<b>2.563</b>	<b>2.78</b>	<b>3.136</b>	<b>2.455</b>	<b>2.165</b>	<b>2.073</b>	<b>2.077</b>
	GGG	0.453	0.591	0.546	0.431	0.662	0.266	0.208
His	CAT	1.518	<b>1.688</b>	1.584	1.55	1.248	1.328	1.337
	CAC	0.777	0.514	0.512	0.567	0.874	0.794	0.755
	TTA	1.452	<b>2.379</b>	<b>1.737</b>	1.477	1.277	1.356	1.341
Leu	TTG	<b>1.883</b>	<b>1.605</b>	<b>1.858</b>	<b>1.963</b>	1.322	1.184	1.171
	CTT	<b>1.753</b>	1.272	<b>1.821</b>	1.47	<b>1.685</b>	<b>1.854</b>	<b>1.881</b>
	CTC	0.516	0.36	0.449	0.403	0.871	1.087	1.104
Ile	CTA	0.563	0.494	0.441	0.558	0.693	0.791	0.795
	CTG	0.362	0.346	0.223	0.551	0.628	0.621	0.601
	ATT	<b>1.799</b>	<b>1.768</b>	<b>2.111</b>	<b>1.708</b>	<b>1.667</b>	1.524	1.509
Lys	ATC	0.46	0.382	0.369	0.371	0.724	0.75	0.75
	ATA	0.788	1.044	0.744	1.021	0.856	1.073	1.068
	AAA	1.092	1.199	1.291	1.039	1.185	1.491	1.502
Phe	AAG	0.909	0.801	0.777	1.002	0.957	0.657	0.639
	TTT	<b>1.635</b>	<b>1.762</b>	<b>1.767</b>	<b>1.698</b>	1.25	1.22	1.22
	TTC	0.553	0.338	0.301	0.376	0.781	0.872	0.866
Pro	CCT	<b>2.294</b>	<b>2.397</b>	<b>2.568</b>	<b>2.057</b>	<b>1.945</b>	<b>2.137</b>	<b>2.151</b>
	CCC	0.897	0.564	0.717	0.633	0.755	0.543	0.519
	CCA	1.573	1.297	1.377	1.518	1.224	<b>1.782</b>	<b>1.785</b>
Ser	CCG	0.569	0.404	0.216	0.463	0.366	1.064	1.06
	TCT	<b>2.235</b>	<b>2.242</b>	<b>2.455</b>	<b>1.842</b>	<b>2.019</b>	<b>1.978</b>	<b>2.003</b>
	TCC	0.538	0.392	0.457	0.724	1.1	0.851	0.851
Thr	TCA	1.307	0.933	1.153	0.963	1.39	<b>1.726</b>	<b>1.662</b>
	TCG	0.411	0.417	0.244	0.36	0.395	0.399	0.389
	AGT	1.494	<b>2.119</b>	<b>2.037</b>	<b>2.067</b>	1.388	1.507	1.503
Tyr	AGC	0.7	0.439	0.527	0.655	0.592	0.662	0.643
	ACT	<b>1.743</b>	<b>2.515</b>	<b>2.242</b>	<b>1.936</b>	<b>2.052</b>	<b>1.892</b>	<b>1.893</b>
	ACC	0.573	0.568	0.541	0.716	0.737	0.577	0.565
Val	ACA	<b>1.606</b>	1.119	1.287	1.302	1.155	<b>1.702</b>	<b>1.714</b>
	ACG	0.324	0.364	0.335	0.314	0.28	0.663	0.658
	TAT	1.461	<b>1.679</b>	1.587	<b>1.628</b>	1.251	1.101	1.091
Val	TAC	0.585	0.435	0.493	0.486	0.782	1.029	1.03
	GTT	<b>2.201</b>	<b>2.791</b>	<b>2.818</b>	<b>2.111</b>	<b>1.84</b>	<b>2.013</b>	<b>2.021</b>
	GTC	0.584	0.346	0.454	0.445	0.819	0.753	0.722
Val	GTA	0.537	0.878	0.496	0.788	0.852	1.141	1.152
	GTG	0.919	0.433	0.455	0.832	0.829	0.665	0.641

bold values represents the over-preferred codons.

than transversions. Transitions acts as silent substitutions because of the wobble and have no observable effect on the organism's phenotype. The transition and transversion mutations may cause a silent mutation if the mRNA codon codes for synonymous codon for the same amino acid, a missense mutation if the mRNA codon generates for a different amino acid, or a nonsense mutation if the mRNA codon produce a stop codon.

#### 2.2.5. Codon adaptation index

The most widespread measure, CAI is used to analyze codon usage bias. Unlike other measures, CAI calculates the deviation of a given

coding sequence regarding a reference set of protein-coding sequences. Relative adaptiveness ( $w_i$ ), for each amino acid in a coding sequence, is computed from a reference coding sequence set.

$$w_i = \frac{f_i}{\max(f_j)} \quad (7)$$

where  $f_i$  is the frequency of the synonymous codon and  $f_j$  indicates the most frequent synonymous codon of particular amino acid.

CAI of a coding gene is expressed as the geometric mean of the weight corresponding to each codon over the gene sequence of length  $L$  (Jansen et al., 2003).



**Table 3**  
The average proportions of slow codons and slow di-codons.

Corona Type	Non-repeated Slow Codon Proportion	Repeated Slow Codon Proportion	Non-overlapped Slow Di-codon Proportion	Overlapped Slow Di-codon Proportion
229E	0.283612	0.220118	0.062016	0.049627
HKU1	0.263711	0.245577	0.07943	0.061318
NL63	0.270957	0.24975	0.089925	0.071582
OC43	0.269984	0.23228	0.067392	0.056349
MERS-CoV	0.279569	0.215358	0.059871	0.049781
SARS-CoV	0.237613	0.176295	0.034371	0.029054
2019-nCoV	0.235117	0.17425	0.033134	0.027704

$$CAI = \left( \prod_{i=1}^L w_i \right)^{\frac{1}{L}} \quad (8)$$

CAI is a quantitative measure used to estimate the level of gene expression by considering the codon usage sequence (Sharp and Li, 1987).

### 2.2.6. Codon position-specific parameters

The parameters considered in coding sequences of different coronaviruses to determine codon bias are the frequency composition of G + C base in first (GC1), second (GC2), and third (GC3) positions in codons and the composition of A + T base at AT1, AT2, AT3 positions.

## 3. Results and discussion

This section provides a detailed analysis to compare various CoVs relating to different parameters that influence codon usage bias and translation rate.

### 3.1. RSCU analysis

The codon usage patterns are peculiar to family, genus and even at the species level of viruses. To analyze this specificity, RSCU values for 59 relevant codons that code for 18 essential amino acids (90%) was determined for all the CDs of varied CoVs. The frequency of each codon type is measured by using a codon usage package of sequence manipulation suite (Stothard, 2000). Three stop codons TAA, TAG, TGA, and two codons ATG, TGG which uniquely code for methionine and tryptophan amino acids respectively are not included in

RSCU analysis. Synonymous codon usage (SCU) bias is a unique trait in many organisms (Wong et al., 2010) and has been noted in various viruses (McBride et al., 2014; Sui et al., 2014; Yang and Chen, 2020). As the coronavirus depends on the host cell factory for its duplication, synonymous codon usage bias plays a crucial role in virion assembly and mRNA translation of the virus.

To analyze SCU bias, RSCU is used and its values of 59 relevant codons that possibly show bias in their usage were examined for different coronavirus as listed in Table 2. Based on the codon usage, the synonymous codon RSCU values are categorized into five groups: i) The RSCU value of 1.0 indicates unbiased codons, it means that particular codon has not shown a sign of bias for that amino acid and codon are chosen equally likely. ii) The synonymous codons whose RSCU value greater than 1.0 indicates that it has positive codon usage bias and are termed as preferred codons. iii) RSCU value less than 1.0 represents that codon has negative usage bias and indicated as less preferred. iv) if the RSCU value is greater than 1.6 the codons are treated as over-preferred, and v) the codons with RSCU values less than 0.6 are treated as under-preferred (Wong et al., 2010).

Identified the average RSCU values of eight codons that belongs to seven amino acids GTT (Val), CCT (Pro), GCT (Ala), TCT (Ser), ACT (Thr), GGT (Gly), AGA (Arg), and CGT (Arg) are over-preferred in all CoVs. The amino acids Ser and Ala are under-preferred with TCG, GCG in all CoVs and amino acids Arg, Gly with CGG, GGG are under-preferred in all except MERS-CoV. In both 2019-nCoV and SARS-CoV, have 13 codons CTT (Lue), GTT (Val), TCT (Ser), TCA (Ser), CCT (Pro), CCA (Pro), ACT (Thr), ACA (Thr), GCT (Ala), CGT (Arg), AGA (Arg), AGG (Arg), and GGT (Gly) are over-preferred, which represent the commonality between these two viruses. The strong deficiency of CpG sites are the consequence of almost all the RSCU values < 0.5 of CGN/NCG codons in all CoVs. To avoid congenital immune responses generally RNA viruses maintain CpG suppression. The 2019-nCoV and SARS-CoV have the fewer number (six) of negatively biased codons than other CoVs 229E, HKU1, NL63, and OC43 that have largely negative biased codons 17, 24, 24, and 16 respectively. The 2019-nCoV has 40 preferred codons (0.6 < average RSCU < 1.6), and 13 over-preferred (average RSCU ≥ 1.6) which may implies a higher protein synthesis rate.

### 3.2. Slow codons and slow di-codons analysis

The average proportions of human slow codons and slow dicodons are extracted from 4143 CDs of seven different CoVs as shown in Table 3. Non-

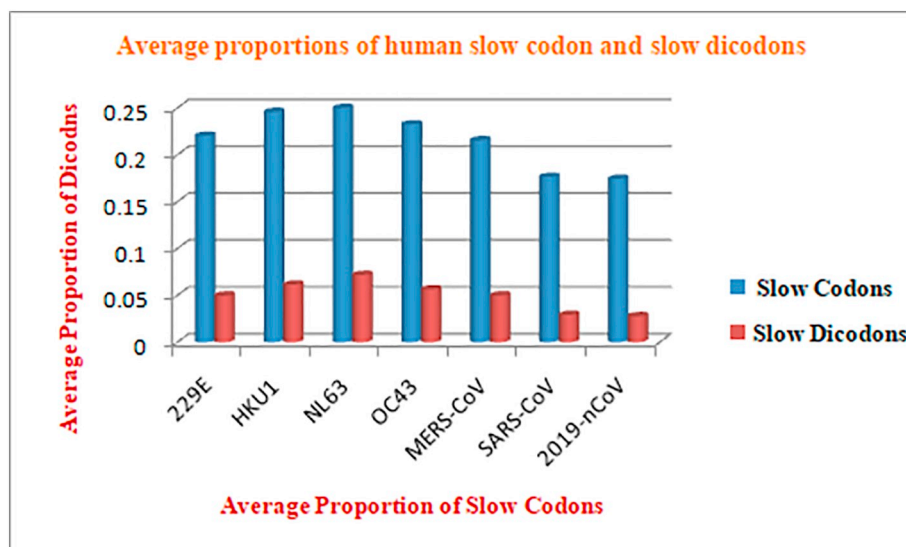


Fig. 1. The average proportions of human slow codons and slow di-codons.

**Table 4**  
Mutations found in various DNA sequences of 2019-nCoV, MERS-CoV, and NL63.

CoV Type	Number of Transition: Transversion Mutations	Silent Mutations	Missense Mutations	Nonsense Mutations	CoVs-Strain
2019-nCoV	1: 1	-	GTA(11082)→CTA, TAC(28143)→CAC	-	CHN/Yunnan-01/2020 MT049951
	5: 0	CTC(18059)→CTT	CCT(17746)→CTT, TAT(17857)→TGT etc.	-	USA/WA3- UW1/ 2020 MT163719
	1: 0	-	CCT(14407)→CTT	-	ZAF/R03006/2020 MT324062
	3: 0	GTT(18125)→GTC	CCT(14407)→CTT, GCT(14785)→GTT	-	GRC/10/2020 MT328032
	2: 2	TAC(14804)→TAT, CGT(17246)→CGC	GTA(11082)→TTA, GGT(26143)→GTT	-	BRA/SP02cc/2020 MT350282
	1: 0	-	CCT(14407)→CTT	-	IND/GBRC1/2020 MT358637
	2: 1	-	AGT(1396)→AAT, GTA(11082)→TTA	-	TWN/CGMH-CGU-05/2020 MT370518
	2: 3	-	AGT(1396)→AAT, GTA(11082)→TTA etc.	-	LKA/COV38/2020 MT371047
MERS-CoV	46: 10	CTT(776)→CTG, CCC(1832)→CCA etc.	CAT(749)→CAG, ATA(1453)→ACA etc.	-	HCoV-EMC MH306207
	43: 5	AGA(3275)→AGG, GTC(12683)→GTT etc.	TGT(541)→TAT, CAT(749)→CAG etc.	-	HCoV-EMC MH013216
	57: 14	CCC(1832)→CCA, AGA(3275)→AGG etc.	CAT(749)→CAG, TCG(1381)→TTG etc.	CAG(13395)→TAG, GAG(23553)→TAG etc.	HCoV-EMC MH454272
	57: 14	CCC(1832)→CCA, CTG(6285)→TTG	CTA(652)→CAA, CAT(749)→CAG etc.	CAG(13395)→TAG, CAA(29850)→TAA	2366 MH432120
	57: 14	CTG(7554)→TTG, CTC(8501)→CTT	CTA(1903)→CCA, TTG(2773)→TCG etc.	GAG(23553)→TAG, CAA(29850)→TAA etc.	2363 MH395139
NL63	48: 6	TGC(12974)→TGT, GCC(13352)→GCT etc.	ATT(17433)→GTT, TCT(17620)→TTT etc.	GAA(20799)→TAA, TTT(414)→CTT,	Haiti-1/2015 KT266906
	67: 12	TGT(14591)→TGC, CTC(14672)→CTT etc.	TAT(2373)→CAT etc. CTC(7740)→TTC,	GAA(20799)→TAA, TTG(21478)→TAG	UF-1/2015 KT381875
	70: 12	GAA(12902)→GAG, TGC(12974)→TGT etc.	CGT(9159)→TGT etc. GAA(12902)→GAG,	GAA(20799)→TAA, TTG(21478)→TAG	UF-2/2015 KU521535
	57: 14	GAA(12902)→GAG, TGC(12974)→TGT	TGC(12974)→TGT etc. AGT(13293)→TGT,	GAA(20799)→TAA, TTG(21478)→TAG	UF-2/2015 KX179500
	21: 46	CTT(16560)→TTG, AAA(16616)→AAG etc.	GAT(14627)→GAA etc.	-	UNKNOWNCS124012 CS124012

**Table 5**  
The average values of different parameters of seven coronaviruses.

Parameter	229E	HKU1	NL63	OC43	MERS-CoV	SARS-CoV	2019-nCoV
ENCs	44.63	39.26	38.36	45.24	49.56	43.75	43.91
GC3s	0.326	0.225	0.241	0.297	0.369	0.326	0.322
GC2s	0.382	0.355	0.366	0.371	0.397	0.365	0.363
GC1s	0.451	0.414	0.455	0.455	0.484	0.456	0.457
AT3	0.674	0.775	0.759	0.703	0.631	0.674	0.678
AT2	0.618	0.645	0.634	0.629	0.603	0.635	0.637
AT1	0.549	0.586	0.545	0.545	0.516	0.544	0.543

repeated slow codon proportion is the ratio between unique slow codons to unique non-slow codons. Repeated slow codon proportion is calculated as in eq. (5). In overlapped slow di-codons, any two consecutive slow codons are paired in a coding sequence and form slow di-codons. Whereas, in non-overlapped slow di-codons, a slow codon that is paired with consecutive slow codon doesn't pair with other immediate right consecutive slow codons. The proportion of slow di-codons is calculated using eq. (6). The order of average proportion of repeated slow codons is 2019-nCoV (0.174) < SARS-CoV (0.176) < MERS-CoV (0.215) < 229E (0.220) < OC43 (0.232) < HKU1 (0.245) < NL63 (0.249) and non-overlapped slow di-codons is 2019-nCoV (0.033) < SARS-CoV (0.034) < MERS-CoV (0.059) < 229E (0.062) < OC43 (0.067) < HKU1 (0.079) < NL63 (0.089) as shown in Fig. 1. The average proportion order of overlapped slow di-codons are 2019-nCoV (0.027) < SARS-CoV (0.029) < 229E (0.049) < MERS-CoV (0.049) < OC43 (0.056) < HKU1 (0.061) < NL63 (0.071). The proportion of slow codons and slow di-codons is inversely proportional to protein synthesis rate (Yang and Chen, 2020). So, the protein translation rates of different CoVs are as shown 2019-nCoV > SARS-CoV > MERS-CoV > 229E > OC43 > HKU1 > NL63.

The number of slow codons and slow di-codons are very less in 2019-nCoV that suggest the protein synthesis rate of 2019-nCoV is more than other types of CoVs. The protein synthesis rate is proportional to the rate of virus replication, which indicates the rapid spread of infection in humans.

### 3.3. Analysis of mutational bias in different CoVs

We performed mutation analysis of three coronavirus (NL63, MERS-CoV, and 2019-nCoV) DNA sequences and examine the transformations of slow codons and non-slow codons due to mutations. A sample of mutations found in the various DNA sequences are shown in Table 4 and the complete set of mutations and its analysis can found at the supplementary material. We identified and analyzed transition, transversion, silent, missense, and nonsense mutations at codon level in CDS that reveals the genetic diversity of various CoVs. The mutation rate in 2019-nCoV is very less compare with MERS-CoV and NL63. In 2019-nCoV, we pointed out silent and missense mutations whereas in other two CoVs nonsense mutations are also recognized. In MERS-CoV and NL63 silent mutations are very high compare with 2019-nCoV. The mutation rates in 2019-nCoV DNA strains of human collected from countries like USA, Greece, Brazil, and Srilanka have higher than the China, India, and South Africa. Due to point mutations at codon level the transformation of slow codons to non-slow codons found to be high that may impact the protein synthesis rate. The results provide evidence for genetic diversity and fast evolution of new corona viruses.

### 3.4. ENC vs. GC3 analysis

The ENC regulate the degree of preference of the codons in the process of decoding. ENC ranges between 20 and 61 codons and inversely correlated with codon bias. High ENC values indicate that the





**Table 7**  
The CAI values for each codon of seven coronaviruses.

Codon	229E	HKU1	NL63	OC43	MERS-CoV	SARS-CoV	2019-nCoV
AAA	1	1	1	1	1	1	1
AAC	0.799	0.601	1	0.728	1	1	1
AAT	1	1	0.829	1	0.985	0.999	0.98
AAG	0.296	0.377	0.245	0.463	0.215	0.315	0.292
ACA	1	1	1	1	1	1	1
ACC	0.8	0.774	0.802	0.629	0.864	0.819	0.85
ACT	0.716	0.667	0.676	0.504	0.886	0.811	0.802
ACG	0.45	0.329	0.399	0.349	0.451	0.24	0.246
ATA	0.353	0.343	0.349	0.403	0.437	0.33	0.307
ATC	0.457	0.36	0.383	0.317	0.456	0.477	0.477
ATT	1	1	1	1	1	1	1
AGA	1	1	1	1	1	1	1
AGC	0.856	0.702	0.791	0.875	0.932	0.864	0.881
AGT	1	1	0.822	0.893	0.92	1	1
AGG	0.64	0.478	0.703	0.597	0.596	0.486	0.488
CAA	1	1	1	1	1	1	1
CAC	0.841	1	0.798	0.983	1	1	1
CAT	1	0.964	1	1	0.991	0.977	0.966
CAG	0.333	0.3	0.213	0.401	0.294	0.393	0.382
CCA	1	1	1	1	1	1	1
CCC	0.707	0.494	0.647	0.702	0.808	0.671	0.717
CCT	0.737	0.661	0.712	0.819	0.864	0.949	0.941
CCG	0.483	0.358	0.349	0.426	0.449	0.325	0.339
CTA	0.407	0.146	0.136	0.22	0.579	0.503	0.493
CTC	0.286	0.118	0.141	0.18	0.496	0.349	0.352
CTT	0.694	0.299	0.454	0.474	1	0.866	0.869
CTG	0.303	0.128	0.167	0.208	0.391	0.37	0.361
CGA	0.321	0.165	0.201	0.183	0.364	0.15	0.152
CGC	0.25	0.066	0.121	0.133	0.305	0.107	0.104
CGT	0.424	0.151	0.149	0.194	0.545	0.166	0.161
CGG	0.225	0.076	0.076	0.132	0.27	0.136	0.135
GAA	1	1	1	1	1	1	1
GAC	1	0.491	0.67	0.654	1	1	1
GAT	0.835	1	1	1	0.735	0.866	0.809
GAG	0.475	0.482	0.445	0.543	0.399	0.49	0.495
GCA	0.975	1	1	1	1	0.885	0.927
GCC	0.666	0.51	0.693	0.687	0.507	0.627	0.681
GCT	1	0.668	0.916	0.917	0.78	1	1
GCG	0.312	0.325	0.585	0.492	0.404	0.283	0.308
GTA	0.354	0.477	0.344	0.511	0.425	0.546	0.544
GTC	0.411	0.308	0.319	0.291	0.474	0.489	0.508
GTT	1	1	1	1	1	1	1
GTG	0.427	0.349	0.333	0.406	0.48	0.562	0.558
GGA	1	0.762	1	1	1	1	1
GGC	0.661	0.811	0.688	0.588	0.924	0.771	0.77
GGT	0.583	1	0.788	0.643	0.753	0.644	0.559
GGG	0.493	0.538	0.679	0.365	0.54	0.632	0.644
TAC	0.753	0.568	0.698	0.755	0.85	0.954	0.956
TAT	1	1	1	1	1	1	1
TCA	0.925	0.958	1	1	0.859	0.819	0.809
TCC	0.437	0.586	0.565	0.569	0.678	0.421	0.426
TCT	0.828	0.787	0.896	0.82	1	0.814	0.8
TCG	0.268	0.354	0.354	0.321	0.31	0.189	0.19
TTA	0.921	1	0.999	1	0.786	1	1
TTC	0.469	0.349	0.422	0.398	0.724	0.573	0.578
TTT	1	1	1	1	1	1	1
TTG	1	0.885	1	0.853	0.72	0.863	0.866
TGC	0.586	0.43	0.457	0.546	0.702	0.77	0.777
TGT	1	1	1	1	1	1	1

The correlation values of HKU1, NL63, SARS-CoV, and 2019-nCoV between ENC and GC3 have a high significance ( $P < .05$ ) disclosing the impact of mutational bias.

### 3.5. Codon position specific analysis

The following parameters are considered in CDs of different CoVs to determine codon biases. The frequencies of G + C and A + T at various positions of all codons are determined and listed in the Table 5. In all CoVs studied in this work, the composition of G + C and A + T frequency percentage order is GC3 < GC2 < GC1 < AT1 < AT2 <

AT3. Hence the CoVs have more AT% than the GC%. The genetic codon composition of seven CoVs is shown in Fig. 3. 2019-nCoV and SARS-CoV has similar codon compositions, and also HKU1, NL63 have similar codon compositions. The codons AAT, AAA, ACT, ATT, GCT and the slow codons GGT, TTT, GAT have the highest codon compositions. The analysis reveals the novel information that the slow codon composition in 229E, HKU1, NL63, OC43, and MERS-CoV are higher than SARS-CoV, 2019-nCoV. Hence, 2019-nCoV and SARS-CoV have a higher translation rate than other CoVs.

### 3.6. Codon adaptation index analysis

CAI for each gene of seven CoVs is measured and the minimum, mean, maximum average CAI values are listed in Table 6. The strong correlation exists between the CAI and the level of gene expression (Sharp and Li, 1987). Gene expression level has a direct impact on the rate of protein evolution in various organisms (Pagan et al., 2012). The order of the average CAI values are 229E < NL63 < HKU1 < 2019-nCoV < OC43 < SARS-CoV < MERS-CoV. The CAI values for each codon of seven CoVs are calculated and listed in Table 7. The codons AAA, ACA, ATT, AGA, CAA, CCA, GAA, GGA, GTT, TAT, TTT, and TGT have highest CAI value for all seven CoVs. 2019-nCoV and SARS-CoV have similar CAI values in which the codons AAC, AGT, CAC, GAC, GCT, TTA shows highest values.

### 3.7. CoVs have distinct evolutionary patterns

We constructed phylogenetic tree to analyze the evolutionary relationships among the seven CoVs using the Maximum Likelihood statistical method with Tamura-Nei model, which is implemented in the Mega-X (Kumar et al., 2018). To determine the robustness of the tree nodes, we performed bootstrap analysis with 500 replicates of dataset. The complete host sequences of seven CoVs were put to phylogenetic analysis and generated individual trees along with combined tree. In case of combine tree, two separate clusters were formed where one indicating human CoVs and other represents zoonotic nature CoVs. The combined tree suggests the diversity in various CoVs as shown in Fig. 4. The individual phylogenetic trees of different CoVs suggest diversified structures as shown in supplementary material. This analysis indicated that 2019-nCoV belongs to beta coronavirus and shares the common ancestor with SARS-CoV.

## 4. Conclusion and future works

The protein synthesis rate depends on the selection of synonymous codons (codon usage bias), translation initiation, tRNA availability and ribosome binding. The accurate calculation of codon usage bias is vital to understand the genetic variations and allows comparison among different CoVs. In this study, we evaluated various measures that manifest the importance of codon bias in translation rate. We assessed seven CoVs with various parameters to determine the correlation among them. Analysis of slow codons and slow di-codons proportions states a relation between viral mRNA genes and viral protein synthesis rate in host cells. We have observed the zoonotic nature CoVs (2019-nCoV, SARS-CoV) have great transmission potential and easily adaptable to host cells than human CoVs. The ENC plot, the correlation values of HKU1, NL63, SARSCoV, and 2019-nCoV have a high significance ( $P < .05$ ) disclosing the impact of mutational bias. The SARS-CoV and 2019-nCoV host sequences clustered far below the standard curve stipulating a high codon bias having a notable correlation with gene expression. Moreover, the mutation rate in 2019-nCoV is very less compare with MERS-CoV and NL63. In 2019-nCoV, we pointed out silent and missense mutations. whereas in other two CoVs nonsense mutations are also recognized. In MERS-CoV and NL63 silent mutations are very high compare with 2019-nCoV. Due to the limitation of data, further investigation and analysis are possible on frame

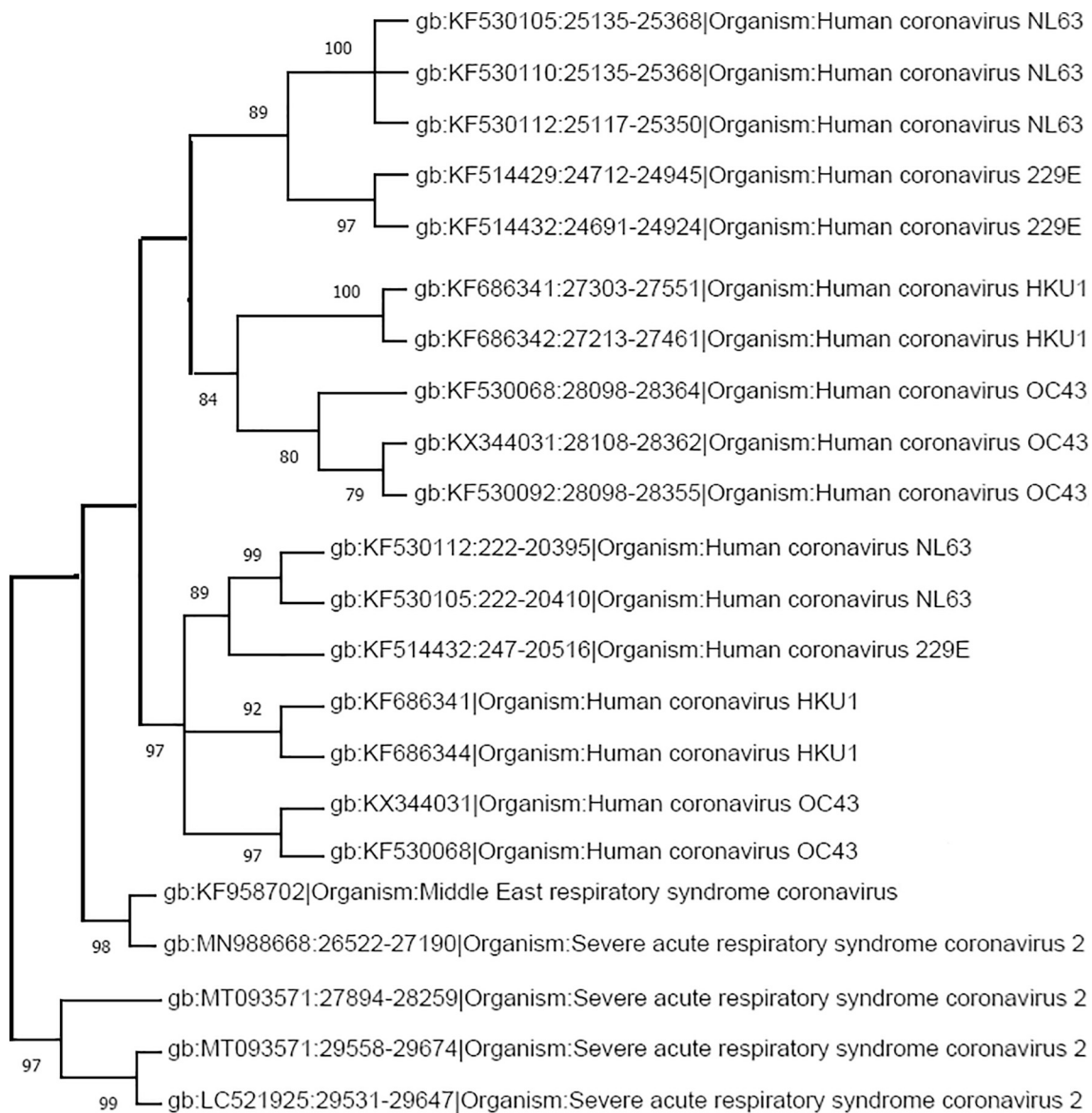


Fig. 4. Phylogenetic tree of human host genome sequences of the representative seven coronaviruses.

shift mutation rates of SARS-COV-2. Further, deep learning techniques can be used to predict patterns related to the novel coronavirus.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.meegid.2020.104432>.

#### References

- Abdul-Rasool, S., Fielding, B.C., 2010. Understanding human coronavirus hcov-nl63. *Open Virol. J.* 4, 76.
- Association, C.P.M., et al., 2020. An update on the epidemiological characteristics of novel coronavirus pneumonia (covid-19). *Chin J Epidemiol* 41 (2), 139–144.
- Brule, C.E., Grayhack, E.J., 2017. Synonymous codons: choose wisely for expression. *Trends Genet.* 33 (4), 283–297.
- Chan, P.P., Lowe, T.M., 2009. Gtrnadb: a database of transfer rna genes detected in genomic sequence. *Nucleic Acids Res.* 37 (Suppl. 1), D93–D97.
- Chekulaeva, M., Landthaler, M., 2016. Eyes on translation. *Mol. Cell* 63 (6), 918–925.
- Chevance, F.F., Hughes, K.T., 2017. Case for the genetic code as a triplet of triplets. *Proc. Natl. Acad. Sci.* 114 (18), 4745–4750.
- Chevance, F.F., Le Guyon, S., Hughes, K.T., 2014. The effects of codon context on in vivo translation speed. *PLoS Genet.* 10 (6).
- Chu, D., Barnes, D.J., Von Der Haar, T., 2011. The role of trna and ribosome competition in coupling the expression of different mrnas in *saccharomyces cerevisiae*. *Nucleic Acids Res.* 39 (15), 6705–6714.
- Dana, A., Tuller, T., 2012. Determinants of translation elongation speed and ribosomal profiling biases in mouse embryonic stem cells. *PLoS Comput. Biol.* 8 (11).
- Gonzalez, D., Giannerini, S., Rosa, R., 2019. On the origin of degeneracy in the genetic code. *Interface Focus* 9 (6), 20190038.
- Hanson, G., Collier, J., 2018. Codon optimality, bias and usage in translation and mrna decay. *Nat. Rev. Mol. Cell Biol.* 19 (1), 20–30.
- Harigaya, Y., Parker, R., 2016. Analysis of the association between codon optimality and mrna stability in *schizosaccharomyces pombe*. *BMC Genomics* 17 (1), 895.
- Health, U.D., Services, H., et al., 2013. Centers for disease control and prevention 2010. The association between school-based physical activity, including physical education, and academic performance.
- Hoffmann, M., Kleine-Weber, H., Krüger, N., Mueller, M. A., Drosten, C., and Pöhlmann, S., 2020. The novel coronavirus 2019 (2019-ncov) uses the sars-coronavirus receptor ace2 and the cellular protease tmprss2 for entry into target cells. *BioRxiv*.
- Husmann, J.A., Patchett, S., Johnson, A., Sawyer, S., Press, W.H., 2015. Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. *PLoS Genet.* 11 (12).
- Iwasaki, S., Ingolia, N.T., 2016. Seeing translation. *Science* 352 (6292), 1391–1392.
- Jansen, R., Bussemaker, H.J., Gerstein, M., 2003. Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Res.* 31 (8), 2242–2251.
- Jenkins, G.M., Holmes, E.C., 2003. The extent of codon usage bias in human rna viruses and its evolutionary origin. *Virus Res.* 92 (1), 1–7.
- Kahn, J.S., McIntosh, K., 2005. History and recent advances in coronavirus discovery. *Pediatr. Infect. Dis. J.* 24 (11), S223–S227.

- Kandeel, M., Ibrahim, A.A., Fayed, M., Al-Nazawi, M., 2020. From sars and mers covs to sars-cov-2: moving toward more biased codon usage in viral structural and non-structural genes. *J. Med. Virol.* 92, 660–666.
- Keller, I., Bensasson, D., Nichols, R.A., 2007. Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes. *PLoS Genet.* 3 (2).
- Koutmou, K.S., Radhakrishnan, A., Green, R., 2015. Synthesis at the speed of codons. *Trends Biochem. Sci.* 40 (12), 717–718.
- Krempf, C., Schultze, B., Herrler, G., 1995. Analysis of cellular receptors for human coronavirus oc43. In: *Corona-and Related Viruses*. Springer, pp. 371–374.
- Kumar, S., Stecher, G., Li, M., Nknyaz, C., Tamura, K., 2018. Mega x: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35 (6), 1547–1549.
- Li, F., 2015. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J. Virol.* 89 (4), 1954–1964.
- Li, F., 2016. Structure, function, and evolution of coronavirus spike proteins. *Ann. Rev. Virol.* 3, 237–261.
- McBride, R., Van Zyl, M., Fielding, B.C., 2014. The coronavirus nucleocapsid is a multi-functional protein. *Viruses* 6 (8), 2991–3018.
- Neuman, B.W., Kiss, G., Kunding, A.H., Bhella, D., Baksh, M.F., Connelly, S., Droese, B., Klaus, J.P., Makino, S., Sawicki, S.G., et al., 2011. A structural analysis of m protein in coronavirus assembly and morphology. *J. Struct. Biol.* 174 (1), 11–22.
- Novoa, E.M., de Poupiana, L.R., 2012. Speeding with control: codon usage, trnas, and ribosomes. *Trends Genet.* 28 (11), 574–581.
- Pagan, I., Holmes, E.C., Simon-Loriere, E., 2012. Level of gene expression is a major determinant of protein evolution in the viral order mononegavirales. *J. Virol.* 86 (9), 5253–5263.
- Pickett, B.E., Sadat, E.L., Zhang, Y., Noronha, J.M., Squires, R.B., Hunt, V., Liu, M., Kumar, S., Zaremba, S., Gu, Z., Zhou, L., Larson, C.N., Dietrich, J., Klem, E.B., Scheuermann, R.H., 2011. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.* 40 (D1), D593–D598.
- Plotkin, J.B., Kudla, G., 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* 12 (1), 32–42.
- Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al., 2015. Codon optimality is a major determinant of mrna stability. *Cell* 160 (6), 1111–1124.
- Radhakrishnan, A., Chen, Y.-H., Martin, S., Alhusaini, N., Green, R., Collier, J., 2016. The dead-box protein dhh1p couples mrna decay and translation by monitoring codon optimality. *Cell* 167 (1), 122–132.
- Ruch, T.R., Machamer, C.E., 2012. The coronavirus e protein: assembly and beyond. *Viruses* 4 (3), 363–382.
- Sharp, P.M., Li, W.-H., 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15 (3), 1281–1295.
- Sharp, P.M., Tuohy, T.M., Mosurski, K.R., 1986. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 14 (13), 5125–5143.
- Sharp, P.M., Emery, L.R., Zeng, K., 2010. Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. B* 365 (1544), 1203–1212.
- Sheikh, A., Al-Taher, A., Al-Nazawi, M., Al-Mubarak, A.I., Kandeel, M., 2020. Analysis of preferred codon usage in the coronavirus n genes and their implications for genome evolution and vaccine design. *J. Virol. Methods* 277, 113806.
- Siddell, S.G., Ziebuhr, J., Snijder, E.J., 2010. Coronaviruses, Toroviruses, and Arteriviruses. *Topley & Wilson's Microbiology and Microbial Infections*.
- Sørensen, M.A., Pedersen, S., 1991. Absolute in vivo translation rates of individual codons in escherichia coli: the two glutamic acid codons gaa and gag are translated with a threefold difference in rate. *J. Mol. Biol.* 222 (2), 265–280.
- Stadler, M., Fire, A., 2011. Wobble base-pairing slows in vivo translation elongation in metazoans. *Rna* 17 (12), 2063–2073.
- Stein, D.A., Skilling, D.E., Iversen, P.L., Smith, A.W., 2001. Inhibition of vesivirus infections in mammalian tissue culture with antisense morpholino oligomers. *Antisense Nucleic Acid Drug Dev.* 11 (5), 317–325.
- Stothard, P., 2000. The sequence manipulation suite: Javascript programs for analyzing and formatting protein and dna sequences. *Biotechniques* 28 (6), 1102–1104.
- Sui, J., Deming, M., Rockx, B., Liddington, R.C., Zhu, Q.K., Baric, R.S., Marasco, W.A., 2014. Effects of human anti-spike protein receptor binding domain antibodies on severe acute respiratory syndrome coronavirus neutralization escape and fitness. *J. Virol.* 88 (23), 13769–13780.
- Sun, X., Yang, Q., Xia, X., 2013. An improved implementation of effective number of codons (n c). *Mol. Biol. Evol.* 30 (1), 191–196.
- Tyrrell, D., Bynoe, M., et al., 1966. Cultivation of viruses from a high proportion of patients with colds. *Lancet* 76–77.
- Wong, E.H., Smith, D.K., Rabadan, R., Peiris, M., Poon, L.L., 2010. Codon usage bias and the evolution of influenza a viruses. Codon usage biases of influenza virus. *BMC Evol. Biol.* 10 (1), 253.
- Wright, F., 1990. The 'effective number of codons' used in a gene. *Gene* 87 (1), 23–29.
- Yampolsky, L.Y., Stoltzfus, A., 2001. Mutational biases. e LS.
- Yang, C.-W., Chen, M.-F., 2020. Composition of Human-Specific Slow Codons and Slow di-Codons in Sars-Cov and 2019-Ncov Are Lower than Other Coronaviruses Suggesting a Faster Protein Synthesis Rate of Sars-Cov and 2019-Ncov. *Journal of Microbiology, Immunology and Infection*.
- Zedalis, J., Eggebrecht, J., 2018. *Biology for AP R Courses*. Rice University.



**Chandra Mohan Dasari** received the B.Tech in Computer Science and Engineering from Kakatiya University in the year 2009, Master of Engineering degree in Computer Science and Engineering from Jadavpur University in the year 2012. He currently works toward Ph.D degree in Computer Science and Engineering from National Institute of Technology, Warangal, Telangana, India. His main research interests are Bioinformatics, Computational Biology, Bigdata Analytics, Machine Learning, and Deep Learning.



**Raju Bhukya** has received his B.Tech in Computer Science and Engineering from Nagarjuna University in the year 2003, M.Tech degree in Computer Science and Engineering (CSE) from Andhra University in the year 2005 and Ph.D in CSE from National Institute of Technology (NIT)Warangal in the year 2014. He is currently working as an Assistant Professor in the Department of CSE in National Institute of Technology, Warangal, Telangana, India. He is currently working in the areas of Bioinformatics and Data Mining.