Journal of Virological Methods 277 (2020) 113806

Contents lists available at ScienceDirect

Journal of Virological Methods

journal homepage: www.elsevier.com/locate/jviromet

Analysis of preferred codon usage in the coronavirus N genes and their implications for genome evolution and vaccine design

Abdullah Sheikh[a,*], Abdulla Al-Taher[b], Mohammed Al-Nazawi[b], Abdullah I. Al-Mubarak[c], Mahmoud Kandeel[b,d]

[a] The Camel Research Center, King Faisal University, Alhofuf, Alahsa 31982, Saudi Arabia
[b] Department of Biomedical Sciences, College of Veterinary Medicine, King Faisal University, Alhofuf, Alahsa 31982, Saudi Arabia
[c] Department of Microbiology, College of Veterinary Medicine, King Faisal University, Alhofuf, Alahsa 31982, Saudi Arabia
[d] Department of Pharmacology, Faculty of Veterinary Medicine, Kafrelsheikh University, Kafrelsheikh 33516, Egypt

ARTICLE INFO

Keywords:
Coronavirus
Nucleocapsid protein
Preferred nucleotides
Amino acid
Codon bias

ABSTRACT

The nucleocapsid (N) protein of a coronavirus plays a crucial role in virus assembly and in its RNA transcription. It is important to characterize a virus at the nucleotide level to discover the virus's genomic sequence variations and similarities relative to other viruses that could have an impact on the functions of its genes and proteins. This entails a comprehensive and comparative analysis of the viral genomes of interest for preferred nucleotides, codon bias, nucleotide changes at the 3rd position (NT3s), synonymous codon usage and relative synonymous codon usage. In this study, the variations in the N proteins among 13 different coronaviruses (CoVs) were analysed at the nucleotide and amino acid levels in an attempt to reveal how these viruses adapt to their hosts relative to their preferred codon usage in the N genes. The results revealed that, overall, eighteen amino acids had different preferred codons and eight of these were over-biased. The N genes had a higher AT% over GC% and the values of their effective number of codons ranged from 40.43 to 53.85, indicating a slight codon bias. Neutrality plots and correlation analyses showed a very high level of GC3s/GC correlation in porcine epidemic diarrhea CoV (pedCoV), followed by Middle East respiratory syndrome-CoV (MERS CoV), porcine delta CoV (dCoV), bat CoV (bCoV) and feline CoV (fCoV) with *r* values 0.81, 0.68, -0.47, 0.98 and 0.58, respectively. These data implied a high rate of evolution of the CoV genomes and a strong influence of mutation on evolutionary selection in the CoV N genes. This type of genetic analysis would be useful for evaluating a virus's host adaptation, evolution and is thus of value to vaccine design strategies.

1. Introduction

Coronaviruses (CoVs) are enveloped, positive-stranded RNA viruses containing a genome of ~30 kb and four structural proteins, namely, spike (S), envelope (E), membrane (M) and nucleocapsid (N) (Siddell et al., 2005). The S protein regulates virus attachment to the receptor of the target host cell (Cavanagh, 1995); the E protein functions to assemble the virions and acts as an ion channel (Ruch and Machamer, 2012); the M protein, along with the E protein, plays a role in virus assembly and is involved in biosynthesis of new virus particles (Neuman et al., 2011); and the N protein forms the ribonucleoprotein complex with the virus RNA (Risco et al., 1996). The N protein is a multifunctional structural protein with distinct characteristics like enhancing transcription of the virus genome, associating with other proteins (M protein) during virion assembly, and inducing toxicity to the host cell by disrupting various cell activities (Berry et al., 2012; McBride et al., 2014). The N protein is the most conserved and stable protein among the CoV structural proteins; whereas, the S protein undergoes several drastic changes during virus infection. For instance, its large parts are cleaved during infection by cellular proteases and expose the receptors to activate viral attachment to the host (Fiscus, 1987; Wu et al., 2004a, 2004b; Maache et al., 2006; Gao et al., 2013). Additionally, the S protein is prone to mutations, especially in the amino acids associated with the spike protein-cell receptor interface, in order to overcome host immunity (Wu and Yan, 2005; Sui et al., 2014). In an interesting study, the N gene of the CoV was found to be more effective for evaluating the codon usage bias than the S gene (Ahn et al., 2009). Studies reported that the N protein produced from prokaryotes has been used to generate specific antibodies against various animal coronaviruses including SARS (Loa et al., 2004; Timani et al., 2004; Wu et al., 2004a, 2004b; Blanchard et al., 2011). The recombinant antigenic N protein from hCoV OC43 used against the rabbit polyclonal

* Corresponding author.

https://doi.org/10.1016/j.jviromet.2019.113806
Received 27 November 2018; Received in revised form 24 November 2019; Accepted 20 December 2019

antibodies specific for hCoV OC43 and did not crossreact with other coronaviruses (SARS CoV and hCoV 229E) (Liang et al., 2013). Moreover, it was tested in different aged human serum samples and exhibited strong reactivity due to the effective central portion (174–300 amino acids) of the N protein followed by C (301–448) & N (1–173) terminal portions (Lee et al., 2008; Yu et al., 2008; Liang et al., 2013). Hence the N protein functions as a sensitive and specific diagnostic tool for hCoV OC43 (Di et al., 2005; He et al., 2005) and it has been further useful in the detection of SARS CoV infection (after the first day of infection) (Che et al., 2004). A similar study on SARS CoV N protein reported immunodominant regions N1 (1–422 amino acids) and N3 (110–422 amino acids) produced specific antigens in BALB/C mice and it reacted with the serum of SARS patients hence it can be used as effective SARS DNA vaccine (Dutta et al., 2008). The N protein of CoV expressed in recombinant raccoon poxvirus revealed an efficient vaccine against feline infectious peritonitis virus infection when administered subcutaneously (Wasmoen et al., 1995).

It is essential to investigate viral gene structures and compositions at the codon or nucleotide level to disclose the mechanisms of virus-host relationships and virus evolution (Bahir et al., 2009; van Hemert et al., 2016). There are 20 amino acids encoded by 61 codons which means that an amino acid could be coded by more than one codon. These alternative codons, up to 6 codons per amino acid, are known as synonymous codons (Nakamura et al., 2000). During gene to protein translation process, some synonymous codons are preferred over others. This is known as codon bias or codon usage bias. Viral genes and genomes exhibit varying numbers of synonymous codons depending on the host (Lloyd and Sharp, 1992). Additionally, codon usage in a virus is influenced by selection pressure and compositional constraints determined by the virus-host system (Karniychuk, 2016). Selective forces act on the gene sequences which maintain the codon bias and gene evolution (Ikemura, 1985; Sharp and Li, 1987; Sharp et al., 1993). Codon bias helpful in the analyzing the horizontal gene transfer as the key evolutionary force to study the molecular evolution of the genes (Doolittle, 1998; Ochman et al., 2000; Woese, 2002). Codon bias occurs during protein expression and it will be same in an organism's genes when there is a similar tRNA content (Kanaya et al., 2001) Codon bias influences the function of the protein and its translation efficiency (Chaney and Clark, 2015; Supek, 2016).

The aim of this study was to carry out a comprehensive analysis of various characteristics, of the N genes of 13 different CoVs, including preferred nucleotides, preferred codons, codon bias, and preferred synonymous codon usage, and to provide an understanding of the codon patterns of these viruses in relation to their hosts and genome evolution.

2. Materials and methods

2.1. Gene data collection and analytical programs

The N genes of 13 different CoV species, viz., Porcine epidemic diarrhea CoV (pedCoV) (171), Middle East respiratory syndrome-CoV (MERS CoV) (265), Infectious bronchitis CoV (ibCoV) (279), Camel alpha CoV (cCoV) (31), Porcine delta CoV (dCoV) (74), Transmissible gastroenteritis CoV (tgCoV) (69), Human CoV 229E (hCoV 229E) (34), Bovine CoV (bvCoV) (49), Bat CoV (bCoV) (34), Human CoV HKU1 (hCoV HKU1) (36), Canine CoV (caCoV) (40), Feline CoV (fCoV) (40) and Human CoV OC43 (hCoV OC43) (112) were used in this study. The coding sequences of the N genes along with their accession numbers were obtained from the GenBank database (Supplementary file). CLC Genomics Workbench 12.0 (QIAGEN, Aarhus, Denmark) (2019) (https://www.qiagenbioinformatics.com/) was used to quantify the nucleotide compositions, A + T % and G + C %. The patterns of codon usage and multivariate statistics were assessed using CodonW 1.4.2 (http://codonw.sourceforge.net//), (Peden, 2000) and the GraphPad prism software was used for correlation analysis.

2.2. Codon usage characterisation

The following parameters of the N gene of each of the CoVs were evaluated to determine the codon bias: the percentage and frequency of each of the four nucleotide bases (A, T, G and C), the G + C base incidences at the starting (GC1) and ending nucleotides (GC3) of the codons, and the number of synonymous codons for each amino acid together with the frequencies of each nucleotide at the 3rd position (A3s, G3s, T3s and C3s).

2.3. Relative synonymous codons usage (RSCU) analysis

RSCU computes the ratios of the expected frequencies of synonymous codon usage by the amino acids against their observed frequencies, assuming that a particular amino acid's synonymous codons were utilized equitably. A value of 1 for a codon in the RSCU table means that the observed frequency of codon usage by the amino acid is equivalent to that of the predictable frequency, or indicating no codon usage bias; whereas, RSCU values of < 1 and > 1 indicate negative and positive codon usage biases, respectively. The formula used to calculate RSCU (Behura and Severson, 2013) is:

$$RSCUij = \frac{Xij}{\frac{i}{ni}\sum_{j=1}^{ni} xji}$$

where X_{ij} denotes an amino acid's observed number of codons used and *ni* stands for the amino acid's overall sum of synonymous codons.

2.4. Analysis of relative dinucleotide frequencies

In a gene, the relative dinucleotide frequencies are determined by calculating the ratios of observed to estimated frequencies of the dinucleotides to determine the codon bias. The formula for the calculating the relative frequency of dinucleotides is:

$$(\mathrm{O/E})\ XpY = [\mathrm{f}\ (XY)/\mathrm{f}\ (X)\mathrm{f}\ (Y)]$$

where f(X) and f(Y) are the single nucleotide frequencies and f(XY) stands for the observed frequency of dinucleotides.

Relative dinucleotide frequency values of < 0.78 denote under-representation of the dinucleotide usage and values of > 1.23 indicate over-representation (Chen and Chen, 2014a). The mentioned values represent the relative abundance of dinucleotides compared to a random distribution.

2.5. Determination of effective number of codons (ENc)

Codon usage bias in a gene can be effectively measured by determining the ENc. ENc values range from 20-61. Higher ENc values indicate low codon bias in which more synonymous codons are used for the amino acids, while lower ENc values represent high codon bias with low numbers of synonymous codons used for the amino acids. Generally, a gene with a strong codon usage bias has an ENc value of 35 or less.

2.6. Assessment of the effect of mutational pressure on codon usage bias

The codon usage bias pattern was analyzed to assess the effective mutational pressure using the ENc plot, in which the GC3 incidence values were plotted against the ENc values (Jenkins and Holmes, 2003; Chen et al., 2004). In ENc plot, the dots represent the individual genes which lie below the curve of expected values subject to mutation pressure. ENc values were interrelated with the mutational pressure which are spread along with the standard curve of GC3 – ENc relation (Fig. 1) (Jenkins and Holmes, 2003; Shi et al., 2016).

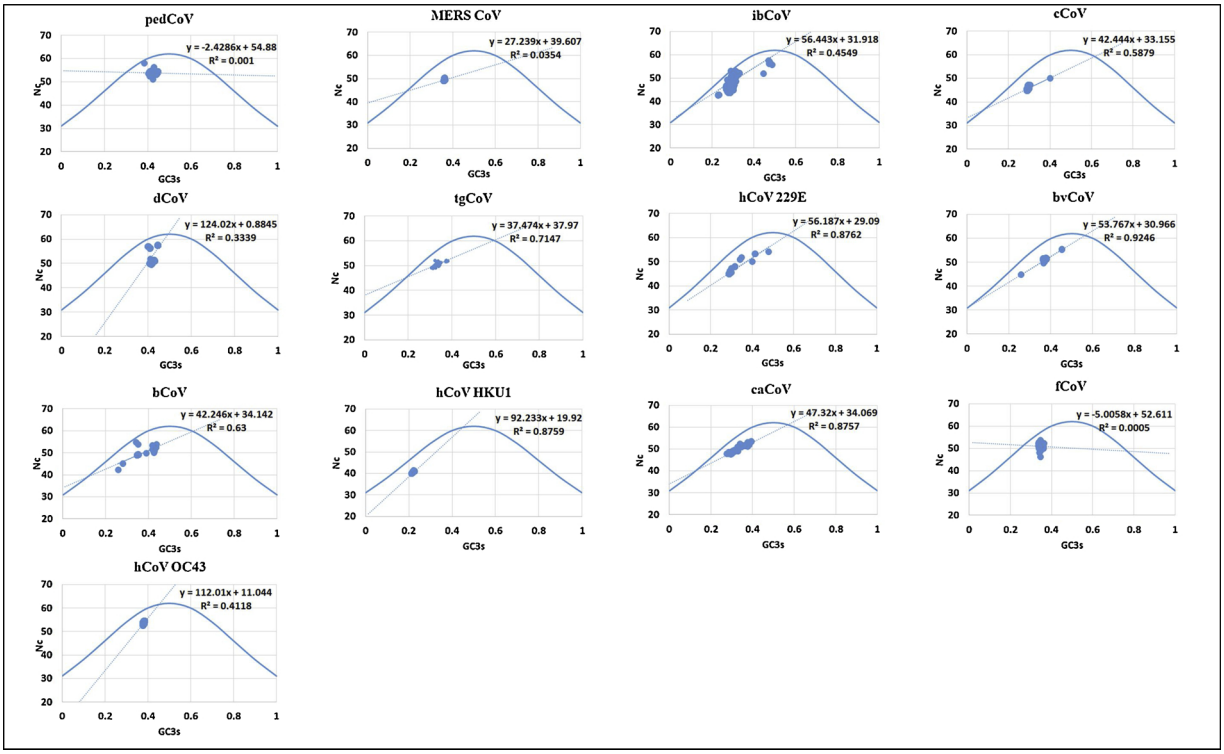

Fig. 1. ENc Plots of N genes from 13 different CoVs representing the relation between GC3s and Nc frequencies.
GC nucleotide frequencies at third positions (GC3s) plotted against the effective number of codons (Nc). GC3s and Nc regression is denoted by a linear dotted line and the solid line represents the relation between GC3s and Nc

2.7. Assessment of the influence of natural-selection on codon usage bias

Neutrality plot analysis was used to evaluate the bias of codon usage as it influenced by natural-selection, the codon adaptation index, and the indices of aromaticity (AROMO) and hydropathicity (GRAVY) (Kumar et al., 2016). It was plotted with GC1, GC2 against GC3. It estimates the neutrality effect of directional mutation pressure in contrast to selection (Sueoka, 1988). The three nucleotide positions of a codon GC1, GC2 and GC3 are the observed GC contents and mostly the GC3 position has the equal number of A/T and G/C nucleotides. There will be variation between GC1, GC2 against GC3 regression values due to directional mutational pressure.

2.8. Multivariate or correspondence analysis (COA)

COA represents the data geometrically by using RSCU values of the genes (Greenacre, 1984). COA was performed on the N genes of the CoVs, using the CodonW analytical program to analyze the RSCU values and to compare the intragene variations of codon usage in the amino acids (Fellenberg et al., 2001; Perrière and Thioulouse, 2002). Each gene displayed as a 59-dimensional vector (59 synonymous codons represented excluding three stop codons, as well as UGG and AUG encoded by single codon) geometrically shows every codon over 59 orthogonal axes and the variation is projected by the axes (Suzuki et al., 2008; D'Andrea et al., 2011).

3. Results

3.1. Nucleotide compositions of the CoV N genes

Comparative analysis and nucleotide compositions of the N genes of 13 different CoVs revealed the nucleotide A (29.61 %) was the most frequent base and the nucleotide frequencies were A > T > G > C (Table 1). Hence, the viruses used more AT% over GC%. Regardless of nucleotide similarities among the CoVs N genes, the nucleotides at the third position (NT3s) of a codon were observed to have variations which contribute to the codon bias and codon pattern differences. The overall NT3s frequencies were T3s > A3s > C3s > G3s. However, it showed some variations when observed individually by summing the NT3s of each gene in the following order of virus (Table 1): tgCoV, fCoV >

Table 1
Nucleotides composition of N gene of 13 CoVs.

Frequencies of nucleotides	pedCoV	MERS CoV	ibCoV	cCoV	dCoV	tgCoV	hCoV 229E	bvCoV	bCoV	hCoV HKU1	caCoV	fCoV	hCoV OC43
Adenine (A)	0.30	0.29	0.30	0.30	0.27	0.32	0.28	0.29	0.28	0.29	0.31	0.31	0.29
Cytosine (C)	0.22	0.25	0.19	0.22	0.25	0.19	0.19	0.21	0.25	0.20	0.18	0.20	0.22
Guanine (G)	0.24	0.21	0.26	0.21	0.21	0.22	0.21	0.24	0.22	0.18	0.20	0.22	0.24
Thymine (T)	0.22	0.23	0.23	0.26	0.25	0.25	0.29	0.24	0.23	0.31	0.29	0.24	0.23
T3s	0.44	0.45	0.48	0.54	0.39	0.43	0.51	0.46	0.45	0.62	0.44	0.47	0.46
C3s	0.29	0.27	0.14	0.20	0.29	0.24	0.20	0.23	0.28	0.15	0.22	0.24	0.24
A3s	0.28	0.32	0.38	0.33	0.30	0.40	0.33	0.31	0.30	0.34	0.39	0.35	0.30
G3s	0.24	0.17	0.23	0.17	0.23	0.19	0.20	0.23	0.18	0.12	0.20	0.20	0.23

Table 2
Various CoVs representing RSCU values.

amino acid	codon	pedCoV	MERS CoV	ibCoV	cCoV	dCoV	tgCoV	hCoV 229E	bvCoV	bCoV	hCoV HKU1	caCoV	fCoV	hCoV OC43
Phe	UUU	0.923	**1.001**	**1.551**	**1.167**	**1.093**	**1.096**	**1.324**	**1.564**	0.83	**1.541**	**1.24**	**1.033**	**1.66**
	UUC	**1.007**	0.999	0.449	0.833	0.899	0.905	0.676	0.436	**1.172**	0.459	0.76	0.968	0.34
Leu	UUA	0.514	0.002	0.280	0.887	0.782	**1.577**	**0.975**	0.509	0.489	**2.371**	**1.37**	0.843	0.78
	UUG	0.872	0.459	0.682	**1.487**	0.632	1.21	1.602	1.802	**1.008**	**0.842**	1.27	0.919	**1.35**
	CUU	**2.103**	**3.702**	**2.171**	**3.013**	**1.668**	**2.001**	**2.23**	**2.357**	**2.646**	1.978	**1.53**	**1.935**	**1.872**
	CUC	**1.441**	**0.688**	0.485	0.189	**1.092**	0.62	0.332	**0.726**	0.481	0.663	0.6	**1.165**	0.752
	CUA	0.29	0.462	**1.221**	0.265	0.823	0.276	0.488	0.098	0.659	0.006	0.53	0.651	0.002
	CUG	0.778	0.674	1.159	0.154	1.012	0.298	0.369	0.509	0.73	0.146	0.7	0.498	1.239
Ile	AUU	**1.791**	**1.861**	**2.087**	**1.383**	**1.36**	1.302	**1.216**	**1.912**	**1.62**	**1.947**	**1.68**	**1.735**	**1.89**
	AUC	**1.159**	**0.91**	0.384	0.31	0.995	0.125	0.601	0.551	**1.203**	**0.761**	0.42	0.232	0.387
	AUA	0.052	0.229	0.527	1.305	0.649	**1.573**	1.181	0.537	0.176	0.292	**0.9**	**1.034**	**0.731**
Val	GUU	**1.613**	**1.413**	**1.349**	**2.015**	**1.542**	**1.492**	**1.76**	**1.43**	**1.464**	**2.594**	**1.5**	**1.515**	**1.15**
	GUC	0.568	0.711	0.519	0.798	0.769	0.514	0.649	0.864	1.416	0.478	0.52	0.694	1.139
	GUA	0.774	1.174	0.861	0.277	0.906	0.751	0.452	0.864	0.65	0.541	0.97	0.597	0.765
	GUG	**1.043**	0.712	1.265	**0.913**	0.779	1.244	**1.145**	0.837	0.469	0.387	1.01	1.193	0.938
Ser	UCU	**2**	**1.9**	**1.52**	**2.33**	**2.27**	**2.1**	**2**	**1.9**	**1.3**	**2.73**	**2**	**2.49**	**1.8**
	UCC	1.05	0.83	0.259	0.6	0.95	0.7	0.6	0.7	0.86	0.58	0.5	0.82	0.7
	UCA	**1.2**	**1.38**	**2**	1.08	**0.98**	**1.3**	1.2	**0.9**	**1.78**	**1.04**	**1.3**	**1.32**	0.7
	UCG	0.22	0.34	0.222	0.16	0.44	0.1	0.2	0.2	0.23	0.36	0.2	0.11	0.3
	AGU	0.69	1.03	1.1	**1.49**	0.64	1.2	**1.5**	1.7	1.01	1.02	1	0.92	1.66
	AGC	0.85	0.51	0.9	0.35	0.65	0.7	0.5	0.6	0.82	0.29	0.9	0.34	**0.84**
Pro	CCU	**1.39**	1.65	**1.36**	**2.28**	**1.47**	**1.6**	**1.7**	1	**2.39**	**2.28**	1.4	**1.91**	**1.39**
	CCC	1.19	**0.6**	0.325	0.54	0.86	0.5	0.6	**1.3**	0.54	0.72	0.6	0.68	1.05
	CCA	1.35	**1.76**	**2**	**1.06**	1.34	**1.6**	**1.4**	1.2	**0.92**	**0.84**	**1.6**	**1.11**	1.03
	CCG	0.07	0	0.308	0.12	0.33	0.3	0.3	0.5	0.15	0.16	0.4	0.31	0.52
Thr	ACU	**2.04**	**2**	**1.45**	**2.15**	**1.6**	1.3	**1.8**	**2**	**1.89**	**2.67**	**1.2**	**1.08**	**1.9**
	ACC	0.25	**1.2**	0.355	0.31	1.23	0.7	0.4	**1**	**1.05**	**0.63**	0.7	0.82	**1.1**
	ACA	**1.37**	0.67	**1.646**	**1.41**	0.82	**1.4**	**1.4**	0.8	0.79	0.54	**1.8**	1.71	0.86
	ACG	0.34	0.13	0.544	0.13	0.35	0.6	0.3	0.2	0.27	0.16	1.13	0.39	0.14
Ala	GCU	**1.28**	**1.93**	**1.25**	**2.27**	**1.44**	1.1	**2**	**1.4**	**2.17**	**3.08**	**1.6**	**2.25**	**1.6**
	GCC	1.11	**0.99**	0.515	0.38	0.77	**1.5**	0.5	1	0.86	0.51	0.9	**1.14**	1.01
	GCA	1.11	0.85	**1.802**	**1.09**	1.21	1.2	**1.2**	1.3	**0.9**	0.41	1.4	0.58	1.13
	GCG	0.5	0.24	0.43	0.26	0.58	0.1	0.3	0.4	0.08	0	0.2	0.03	0.26

(*continued on next page*)

Table 2 (*continued*)

Tyr	UAU	0.951	0.206	**1.3**	**1.469**	**1.162**	**1.094**	**1.426**	**1.1**	0.506	**1.749**	**1.117**	0.524	**1.03**
	UAC	**1.049**	**1.794**	0.7	0.531	0.841	0.906	0.574	0.9	**1.494**	0.251	0.883	**1.477**	0.97
His	CAU	**1.347**	**1.005**	**1.3**	0.998	**1.317**	0.878	**1.137**	**1.02**	0.97	**1.328**	**1.11**	0.626	**1.469**
	CAC	0.653	0.994	0.7	**1.002**	0.688	**1.122**	0.863	0.98	**1.03**	0.672	0.89	**1.374**	0.53
Gln	CAA	0.469	**1.33**	0.9	**1.154**	0.921	**1.241**	**1.225**	0.89	0.808	**1.516**	**1.359**	**1.059**	0.897
	CAG	**1.531**	0.67	**1.1**	0.846	**1.079**	0.759	0.775	**1.11**	**1.192**	0.484	0.642	0.942	**1.103**
Asn	AAU	**1.209**	**1.06**	**1.6**	**1.231**	0.885	**1.077**	**1.123**	**1.68**	**1.099**	**1.722**	**1.193**	**1.134**	**1.649**
	AAC	0.791	0.94	0.4	0.767	**1.115**	0.923	0.877	0.32	0.901	0.278	0.808	0.866	0.351
Lys	AAA	0.822	**1.032**	0.8	**1.086**	**1.036**	**1.285**	**1.14**	0.78	0.981	**1.52**	**1.147**	**1.078**	0.879
	AAG	**1.178**	0.968	**1.2**	0.914	0.964	0.715	0.86	**1.22**	**1.019**	0.48	0.881	0.923	**1.121**
Asp	GAU	**1.049**	**1.498**	**1.5**	**1.075**	0.928	**1.307**	**1.047**	**1.18**	**1.208**	**1.43**	**1.272**	**1.642**	**1.09**
	GAC	0.951	0.502	0.5	0.925	**1.072**	0.693	0.952	0.82	0.792	0.57	0.715	0.359	0.911
Glu	GAA	**1.234**	0.921	**1.3**	**1.537**	0.894	**1.51**	**1.399**	**1.03**	**1.095**	**1.322**	**1.382**	**1.181**	**1.006**
	GAG	0.766	**1.079**	0.7	0.463	**1.106**	0.49	0.601	0.97	0.905	0.678	0.618	0.819	0.994
Cys	UGU	0.11	**0.02**	**1.7**	**1.96**	0.5	**1.8**	**1.6**	**1**	0.06	**2**	**1.5**	0.66	**1**
	UGC	**1.89**	**0.02**	0.3	0.04	**0.6**	0.2	0.4	**1**	**0.94**	0	0.5	**1.29**	**1**
Arg	CGU	**2.16**	1.15	1.7	**2.31**	**1.32**	**1.4**	1.8	**1.2**	**1.33**	**1.48**	**1.2**	**1.53**	**1.07**
	CGC	1.05	**1.37**	0.9	0.61	0.84	0.8	0.5	0.6	1.2	0.23	0.6	0.58	0.8
	CGA	0.17	0.46	0.3	0.1	0.43	0.1	0.2	0.3	0.6	0.52	0.3	0.18	0.21
	CGG	0.17	0.91	0.1	**0.62**	0.54	0.4	0.6	0.2	0.34	0.36	0.2	0.39	0.38
	AGA	1.21	**1.85**	**2**	1.89	**2.02**	**2.2**	**2**	**2.7**	**2.06**	**2.3**	**2.5**	**2.1**	**2.52**
	AGG	**1.24**	0.23	**1**	0.47	0.87	1.1	**0.9**	0.9	0.47	1.11	0.7	1.22	1.02
Gly	GGU	**1.57**	1.16	**2**	**2.33**	**1.33**	**2**	**2.1**	1.2	1.41	**2.15**	**2**	**2.08**	1.2
	GGC	1.07	0.74	0.3	**1.19**	**1.33**	0.7	**1**	0.8	0.69	0.51	0.6	0.42	0.96
	GGA	1.18	**1.37**	**1.3**	0.47	0.97	**1.1**	0.7	**1.5**	**1.45**	**1.11**	**1.1**	**1.45**	**1.55**
	GGG	0.18	0.74	0.4	0	0.37	0.3	0.2	0.4	0.45	0.24	0.3	0.05	0.3

The values in bold are preferred codons for respective amino acids. The cells with negative biased values have a diagonal line. Over biased codon values are displayed in bold with shaded cells.

pedCoV, caCoV > cCoV, hCoV 229E > ibCoV, bvCoV, hCoV HKU1, hCoV OC43 > MERS CoV, dCoV, bCoV. In NT3s, the T3s nucleotide was the most recurrent one with a frequency of 0.62 and the least recurrent was G3s with a frequency of 0.12 (hCoV HKU1) (Table 1).

3.2. RSCU analysis

Codons with RSCU values were categorized into 3 groups: i) RSCU values < 0.6 denote underrepresented codons (negatively biased); ii) values ranging from 0.6 to 1.6 constitute represented codons (with no bias); and iii) the values > 1.6 indicate over-represented codons (positively biased). A3s and T3s were the most recurrent nucleotides in the represented (preferred) codons and C3s and G3s were the least frequent in overall studied viruses (Table 2). Eighteen amino acids (90 %) were observed with varied codon preferences (Phe, Leu, Ile, Val, Ser, Pro, Thr, Ala, Tyr, His, Gln, Asn, Lys, Asp, Glu, Cys, Arg, Gly) (Table 2). There were eight overrepresented amino acids (Leu, Ser, Pro, Thr, Ala, Tyr, Cys, Arg). Their corresponding codon values are represented in bold with shaded cells in Table 2.

The amino acid Leu overbiased with CUU codon in all the genes except in hCoV HKU1 where it overbiased with UUA. The amino acid Ser overrepresented with UCU in all except UCA in ibCoV and bCoV. The overbiased codon for the amino acid Pro was CCA in MERS CoV and IbCoV, while in others it was CCU. The amino acid Thr over preferred ACU codon in all the genes except ibCoV and caCoV where its preferred ACA. The amino acid Ala highly overbiased with GCU codon in 12 genes while it favored ACA in IbCoV. Similarly, the amino acid Tyr encoded with UAC in MERS CoV while it overrepresented UAU in hCoV HKU1; Cys amino acid overbiased with UGC in pedCoV but UGU in other genes; Arg amino acid preferred CGU codon in pedCoV and CCoV while AGA dominated in others. Overall, among the NT3s of overrepresented or over biased codons, A3s and T3s dominated over C3s and G3s while in negative biased or underrepresented NT3s the order was G3s > C3s > A3s > T3s.

Table 3
Codon Usage Indices of various CoVs.

	pedCoV	MERS CoV	ibCoV	cCoV	dCoV	tgCoV	hCoV 229E	bvCoV	bCoV	hCoV HKU1	caCoV	fCoV	hCoV OC43
ENc	53.85	49.53	48.84	45.84	53.6	50.74	46.95	51.14	50.32	40.43	49.82	50.86	53.85
GC3s	0.42	0.36	0.29	0.29	0.42	0.34	0.31	0.37	0.38	0.22	0.33	0.34	0.38
GC	0.46	0.47	0.45	0.43	0.47	0.41	0.42	0.46	0.48	0.39	0.40	0.43	0.46
GRAVY	−1.07	−0.86	−1.01	−0.89	−0.45	−0.87	−0.57	−0.82	−0.84	−0.84	−0.43	−1.02	−0.86
AROMO	0.06	0.07	0.07	0.06	0.08	0.08	0.07	0.08	0.07	0.10	0.10	0.08	0.08

3.3. ENc and ENc plot

Generally, the values of ENc fall in-between 20–61. As the codon number decreases for a particular amino acid, it results in decrease of ENc value indicating higher codon bias. Conversely, increase in codon number corresponds with less or little codon bias for an amino acid. The ENc values for all the studied CoVs ranged from 40.43 to 53.85 (Table 3). Generally, the estimated average ENc values of RNA viruses span from 38.9 to 58.3 (Jenkins and Holmes, 2003). High ENc values suggest that the CoVs genes are highly conserved along with effective replication, whereas the lowest ENc value i.e. 20 reflects codon usage with extreme bias (one amino acid is coded by a single codon). Our study observed 18 amino acids having different synonymous codons. Furthermore, the RNA viruses usually consist of high ENc values which help it in replication and host adaption with preferred codons.

An ENc plot is useful in analyzing mutational pressure and compositional constraints on codon usage and the compositional bias denoted by the points on the standard curve of ENc and GCs relation. Other forces influencing mutational bias are defined by the points beneath the standard curve. In all the viruses, all points were located below the standard curve, hence suggesting that codon usage bias was influenced by compositional constraints and other factors like virus host interactions and natural selection may influence codon bias.

Correlation values ranged from r = 0.0005 (fCoV) to 0.924 (bvCoV) as shown in Fig. 1. Significance levels were attained in various correlation analyses presented in a supplementary data file. The correlation between GC3s and ENc of ibCoV, tgCoV and bvCoV had a high significance ($P < 0.001$) revealing mutational bias influence along with extra codon usage bias. Whereas the rest of the CoVs did not yield significant correlations reflecting less influence of compositional constraints.

3.4. Neutrality plot

Neutrality plot analyze the neutrality of evolution by evaluating the impact of selection and mutation on codon usage bias. The significant correlation of GC3s and GC1,2 s was achieved through random selection when the genes lie on the slope of unity and then the particular gene is said to be under neutral mutation. The directional mutation pressure on codon usage occurs as the slope moves towards the x-axis. GC1, 2 s were plotted against GC3s and the slope implies the motion at which the mutation and selection forces evolved. The coefficient of regression denotes the equilibrium coefficient of mutation selection as shown in Fig. 2.

The correlation analyses showed a very high correlation in pedCoV followed by MERS CoV, dCoV, bCoV and fCoV with *r* values 0.81, 0.68, -0.47, 0.98, and 0.58 respectively. The used N gene sequences for pedCoV were obtained from a broad range of timescale covering from March 2007 to August 2017. In contrast, MERS CoV N gene sequences were from July 2014 to January 2018. This might have reflected on the rate of compositional variations and adaptation to the host. Therefore, pedCoV showed more attempts to adapt the host by changing its genomic composition. Since MERS CoV is still considered as the emerging virus, its genomic composition is still evolving and its host adaptation is still a matter of debate. Three of the viruses showed moderate or medium correlation i.e. cCoV, hCoV 229E and caCoV with the following *r* values: 0.35, -0.41 and -0.5 (negative correlation). The slopes of regression ranged from -0.8954 to 0.5891 in all the studied viruses represented in Fig. 2. Therefore, this reveals the directional mutational pressure and neutrality influenced them. The supplementary data file includes the AROMO and GRAVY analyses which revealed moderate correlations among the studied CoVs N genes and their varying significance levels likely due to ENc, GC3s, GC variations. Thus, we can infer the codon usage influences from aromaticity and hydropathicity.

4. Discussion

Computational approaches are linked with most of the research studies including genomic analyses, evolution and drug discovery etc. (Kandeel et al., 2009a, b; Kandeel et al., 2009c). In the present work we assessed N gene of different CoVs with various factors such as natural selection, mutational selection and others to determine the codon bias and codon usage indices which regulate virion assembly and transcription of viral RNA in CoVs. The nucleotide contents revealed higher AT% and low GC% as it is common in RNA viruses such as Severe Acute Respiratory Syndrome (SARS) (Jenkins and Holmes, 2003; Gu et al., 2004; Zhou et al., 2005). The ENc values > 35 indicates the slight codon bias due to mutation pressure or nucleotide compositional constraints. This suggests that the RNA viruses with high ENc values adapt to the host with various preferred codons (Jenkins and Holmes, 2003). The positively biased or represented codons of the present study are similar to the two other studies on MERS CoV proteases and pandemic influenza virus (H1N1 and in H3N2) (Kumar et al., 2016; Kandeel and Altaher, 2017). In zika virus and tembusu virus codon usage was driven by the mutation bias (Cristina et al., 2015; Zhou et al., 2015) while in Parvoviridae and pedCoV it was dominated by selection pressure (Shi et al., 2013; Chen et al., 2014b). Some of the viruses observed with the codon bias related to their hosts during their adaptation (Chantawannakul and Cutler, 2008; Bahir et al., 2009; Cheng et al., 2012; Kattoor et al., 2015; Ma et al., 2015; Nasrullah et al., 2015). Studies directed at the conserved regions of viral proteins are useful for developing diagnostic reagents and probes for detecting a range of viruses and isolates in one test and for vaccine development (Du et al., 2010; Johnson et al., 2019). In view of the lower mutation rates and relatively conserved sequences of the coronavirus (CoV) N gene, it would be ideal for studying these genes as an intermediary step in the development of vaccines and diagnostics for these viruses. The present study aids in the understanding of different factors influencing the variations of the N gene among various CoVs and their relationships with their hosts.

Author statement

Dear Dr. Paul

Our manuscript titled “Genomic variations in the N-gene of various Coronaviruses” assess nucleotide and aminoacid variations among the 13 different Coronaviruses and the authors read and approved the final version of the manuscript. All the authors have been contributed in this manuscript.

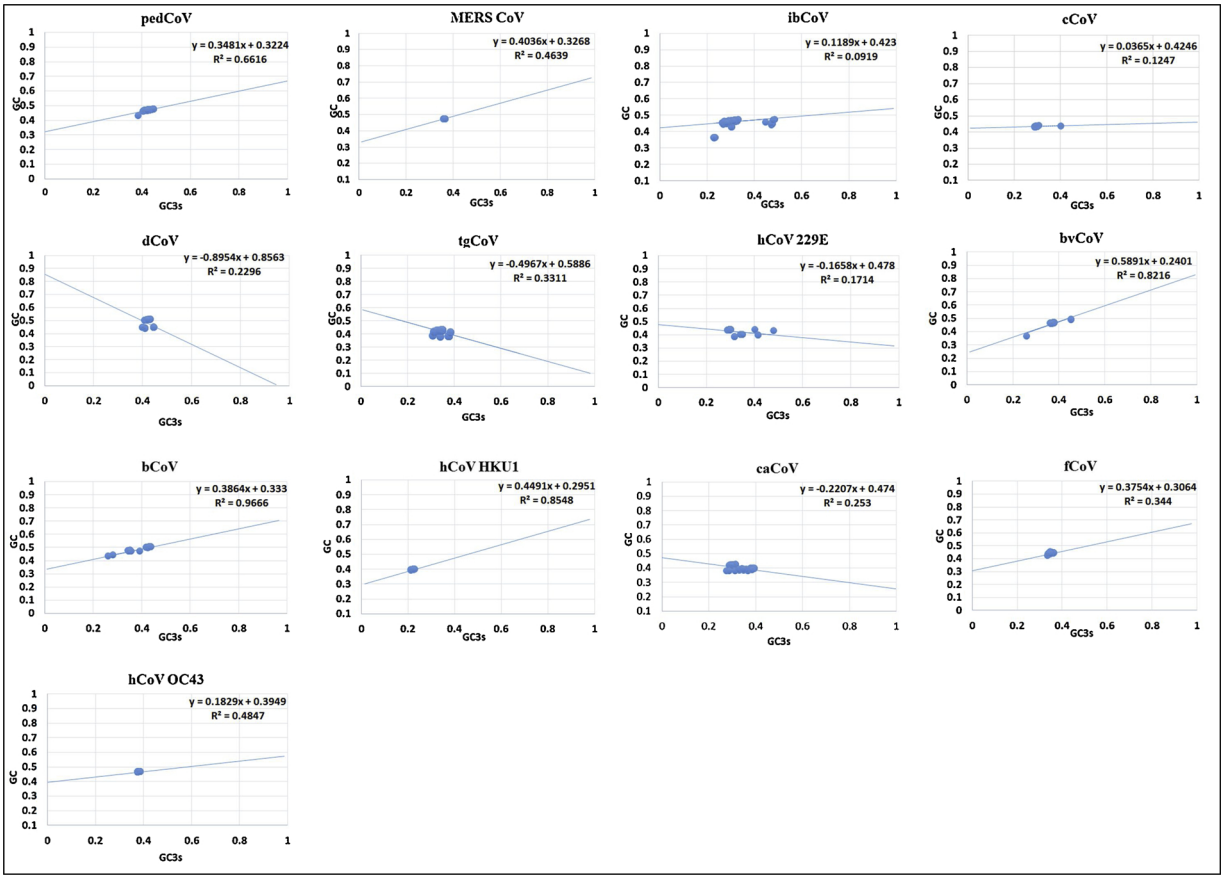

Fig. 2. Neutrality Plots of N genes from 13 different CoVs.
The GC nucleotide base frequencies at the third positions (GC3s) were plotted against the GC frequencies of first and second positions (GC)

I would like to inform you regarding the addition of another author who helped in the manuscript revision for which I had emailed you and there is a slight change in affiliation. The genbank accessions were included for all the viruses in the xl file as a supplementary data. Thanking you for your time and support.

CRediT authorship contribution statement

Abdullah Sheikh: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Writing - review & editing. **Abdulla Al-Taher:** Visualization, Investigation. **Mohammed Al-Nazawi:** Supervision. **Abdullah I. Al-Mubarak:** Writing - review & editing. **Mahmoud Kandeel:** Conceptualization, Methodology, Software, Visualization, Investigation, Validation.

Declaration of Competing Interest

None.

Acknowledgment

The authors acknowledge the Deanship of Scientific Research at King Faisal University for the financial support under strategic projects track [grant number 171001].

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jviromet.2019.113806.

References

Ahn, I., Jeong, B.J., Son, H.S., 2009. Comparative study of synonymous codon usage variations between the nucleocapsid and spike genes of coronavirus, and C-type lectin domain genes of human and mouse. Exp. Mol. Med. 41, 746.
Bahir, I., Fromer, M., Prat, Y., Linial, M., 2009. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. Mol. Syst. Biol. 5, 311.
Behura, S.K., Severson, D.W., 2013. Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. Biol. Rev. 88, 49–61.
Berry, M., Manasse, T.-L., Tan, Y.-J., Fielding, B.C., 2012. Characterisation of human coronavirus-NL63 nucleocapsid protein. Afr. J. Biotechnol. 11, 13962–13968.
Blanchard, E.G., Miao, C., Haupt, T.E., Anderson, L.J., Haynes, L.M., 2011. Development of a recombinant truncated nucleocapsid protein based immunoassay for detection of antibodies against human coronavirus OC43. J. Virol. Methods 177, 100–106.
Cavanagh, D., 1995. The Coronavirus Surface Glycoprotein, the Coronaviridae. Springer, pp. 73–113.
Chaney, J.L., Clark, P.L., 2015. Roles for synonymous codon usage in protein biogenesis. Annu. Rev. Biophys. 44, 143–166.
Chantawannakul, P., Cutler, R.W., 2008. Convergent host-parasite codon usage between honeybee and bee associated viral genomes. J. Invertebr. Pathol. 98, 206–210.
Che, X.Y., Hao, W., Wang, Y., Di, B., Yin, K., Xu, Y.C., Feng, C.S., Wan, Z.Y., Cheng, V.C., Yuen, K.Y., 2004. Nucleocapsid protein as early diagnostic marker for SARS. Emerg. Infect. Dis. 10, 1947–1949.
Chen, S.L., Lee, W., Hottes, A.K., Shapiro, L., McAdams, H.H., 2004. Codon usage between genomes is constrained by genome-wide mutational processes. Proc. Natl. Acad. Sci. 101, 3480–3485.
Chen, Y., Chen, Y.-F., 2014a. Analysis of synonymous codon usage patterns in duck hepatitis A virus: a comparison on the roles of mutual pressure and natural selection. Virusdisease 25, 285–293.
Chen, Y., Shi, Y., Deng, H., Gu, T., Xu, J., Ou, J., Jiang, Z., Jiao, Y., Zou, T., Wang, C., 2014b. Characterization of the porcine epidemic diarrhea virus codon usage bias. Infect. Genet. Evol. 28, 95–100.
Cheng, X.F., Wu, X.Y., Wang, H.Z., Sun, Y.Q., Qian, Y.S., Luo, L., 2012. High codon adaptation in citrus tristeza virus to its citrus host. Virol. J. 9, 113.
CLC Genomics Workbench 12.0 (QIAGEN, Aarhus, Denmark). https://www.qiagenbioinformatics.com/. (Accessed 20 January 2018).
Cristina, J., Moreno, P., Moratorio, G., Musto, H., 2015. Genome-wide analysis of codon usage bias in Ebolavirus. Virus Res. 196, 87–93.

D'Andrea, L., Pintó, R.M., Bosch, A., Musto, H., Cristina, J., 2011. A detailed comparative analysis on the overall codon usage patterns in hepatitis A virus. Virus Res. 157, 19–24.
Di, B., Hao, W., Gao, Y., Wang, M., Wang, Y.D., Qiu, L.W., Wen, K., Zhou, D.H., Wu, X.W., Lu, E.J., Liao, Z.Y., Mei, Y.B., Zheng, B.J., Che, X.Y., 2005. Monoclonal antibody-based antigen capture enzyme-linked immunosorbent assay reveals high sensitivity of the nucleocapsid protein in acute-phase sera of severe acute respiratory syndrome patients. Clin. Diagn. Lab. Immunol. 12, 135–140.
Doolittle, W.F., 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet. 14, 307–311.
Du, L., Zhou, Y., Jiang, S., 2010. Research and development of universal influenza vaccines. Microbes Infect. 12, 280–286.
Dutta, N.K., Mazumdar, K., Lee, B.H., Baek, M.W., Kim, D.J., Na, Y.R., Park, S.H., Lee, H.K., Kariwa, H., Park, J.H., 2008. Search for potential target site of nucleocapsid gene for the design of an epitope-based SARS DNA vaccine. Immunol. Lett. 118, 65–71.
Fellenberg, K., Hauser, N.C., Brors, B., Neutzner, A., Hoheisel, J.D., Vingron, M., 2001. Correspondence analysis applied to microarray data. Proc. Natl. Acad. Sci. 98, 10781–10786.
Gao, J., Lu, G., Qi, J., Li, Y., Wu, Y., Deng, Y., Geng, H., Li, H., Wang, Q., Xiao, H., 2013. Structure of the fusion core and inhibition of fusion by a heptad repeat peptide derived from the S protein of Middle East respiratory syndrome coronavirus. J. Virol. 87, 13134–13140.
Greenacre, M.J., 1984. Theory and Applications of Correspondence Analysis. 1984. Academic Press, London.
Gu, W., Zhou, T., Ma, J., Sun, X., Lu, Z., 2004. Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. Virus Res. 101, 155–161.
He, Q., Du, Q., Lau, S., Manopo, I., Lu, L., Chan, S.W., Fenner, B.J., Kwang, J., 2005. Characterization of monoclonal antibody against SARS coronavirus nucleocapsid antigen and development of an antigen capture ELISA. J. Virol. Methods 127, 46–53.
Ikemura, T., 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2, 13–34.
Jenkins, G.M., Holmes, E.C., 2003. The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res. 92, 1–7.
Johnson, R.F., Schnell, M., Hensley, L.E., Wirblich, C., Coleman, C.M. and Frieman, M.B. 2019. Multivalent vaccines for rabies virus and coronoviruses, U.S. Patent Application 16/091,005.
Kanaya, S., Kinouchi, M., Abe, T., Kudo, Y., Yamada, Y., Nishi, T., Mori, H., Ikemura, T., 2001. Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the E. Coli O157 genome. Gene 276, 89–99.
Kandeel, M., Altaher, A., 2017. Synonymous and biased codon usage by MERS CoV papain-like and 3CL-proteases. Biol. Pharm. Bull. 40, 1086–1091.
Kandeel, M., Kato, A., Kitamura, Y., Kitade, Y., 2009a. Thymidylate kinase: the lost chemotherapeutic target. Nucleic Acids Symposium Series 53. pp. 283–284.
Kandeel, M., Kitamura, Y., Kitade, Y., 2009b. The exceptional properties of plasmodium deoxyguanylate pathways as a potential area for metabolic and drug discovery studies. Nucleic Acids Symposium Series 53. pp. 39–40.
Kandeel, M., Miyamoto, T., Kitade, Y., 2009c. Bioinformatics, enzymologic properties, and comprehensive tracking of Plasmodium falciparum nucleoside diphosphate kinase. Biol. Pharm. Bull. 32, 1321–1327.
Karniychuk, U.U., 2016. Analysis of the synonymous codon usage bias in recently emerged enterovirus D68 strains. Virus Res. 223, 73–79.
Kattoor, J.J., Malik, Y.S., Sasidharan, A., Rajan, V.M., Dhama, K., Ghosh, S., Banyai, K., Kobayashi, N., Singh, R.K., 2015. Analysis of codon usage pattern evolution in avian rotaviruses and their preferred host. Infect. Genet. Evol. 34, 17–25.
Kumar, N., Bera, B.C., Greenbaum, B.D., Bhatia, S., Sood, R., Selvaraj, P., Anand, T., Tripathi, B.N., Virmani, N., 2016. Revelation of influencing factors in overall codon usage bias of equine influenza viruses. PLoS One 11, e0154376.
Lee, H.K., Lee, B.H., Dutta, N.K., Seok, S.H., Baek, M.W., Lee, H.Y., Kim, D.J., Na, Y.R., Noh, K.J., Park, S.H., Kariwa, H., Nakauchi, M., Maile, Q., Heo, S.J., Park, J.H., 2008. Detection of antibodies against SARS-Coronavirus using recombinant truncated nucleocapsid proteins by ELISA. J. Microbiol. Biotechnol. 18, 1717–1721.
Liang, F.Y., Lin, L.C., Ying, T.H., Yao, C.W., Tang, T.K., Chen, Y.W., Hou, M.H., 2013. Immunoreactivity characterisation of the three structural regions of the human coronavirus OC43 nucleocapsid protein by Western blot: implications for the diagnosis of coronavirus infection. J. Virol. Methods 187 (2), 413–420.
Lloyd, A.T., Sharp, P.M., 1992. Evolution of codon usage patterns: the extent and nature of divergence between Candida albicans and Saccharomyces cerevisiae. Nucleic Acids Res. 20, 5289–5295.
Loa, C.C., Lin, T.L., Wu, C.C., Bryan, T.A., Hooper, T., Schrader, D., 2004. Expression and purification of turkey coronavirus nucleocapsid protein in Escherichia coli. J. Virol. Methods 116, 161–167.
Ma, Y.P., Liu, Z.X., Hao, L., Ma, J.Y., Liang, Z.L., Li, Y.G., Ke, H., 2015. Analysing codon usage bias of cyprinid herpesvirus 3 and adaptation of this virus to the hosts. J. Fish Dis. 38, 665–673.
Maache, M., Komurian-Pradel, F., Rajoharison, A., Perret, M., Berland, J.-L., Pouzol, S., Bagnaud, A., Duverger, B., Xu, J., Osuna, A., 2006. False-positive results in a recombinant severe acute respiratory syndrome-associated coronavirus (SARS-CoV) nucleocapsid-based western blot assay were rectified by the use of two subunits (S1 and S2) of spike for detection of antibody to SARS-CoV. Clin. Vaccine Immunol. 13, 409–414.
McBride, R., van Zyl, M., Fielding, B.C., 2014. The coronavirus nucleocapsid is a multifunctional protein. Viruses 6, 2991–3018.
Nakamura, Y., Gojobori, T., Ikemura, T., 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res. 28 292-292.
Nasrullah, I., Butt, A.M., Tahir, S., Idrees, M., Tong, Y., 2015. Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. BMC Evol. Biol. 15, 174.
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., 2011. A structural analysis of M protein in coronavirus assembly and morphology. J. Struct. Biol. 174, 11–22.
Ochman, H., Lawrence, J.G., Groisman, E.A., 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405, 299.
Peden, J.F., 2000. Analysis of Codon Usage. University of Nottingham, pp. 2000. (Accessed 20 January 2018. http://codonw.sourceforge.net/).
Perrière, G., Thioulouse, J., 2002. Use and misuse of correspondence analysis in codon usage studies. Nucleic Acids Res. 30, 4548–4555.
Risco, C., Antón, I.M., Enjuanes, L., Carrascosa, J.L., 1996. The transmissible gastroenteritis coronavirus contains a spherical core shell consisting of M and N proteins. J. Virol. 70, 4773–4777.
Ruch, T., Machamer, C., 2012. The coronavirus E protein: assembly and beyond. Viruses 4, 363–382.
Sharp, P.M., Li, W.H., 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. 4, 222–230.
Sharp, P.M., Stenico, M., Peden, J.F., Lloyd, A.T., 1993. Codon usage: mutational bias, translational selection, or both. Biochem. Soc. Trans. 21, 835–841.
Shi, S.L., Jiang, Y.R., Liu, Y.Q., Xia, R.X., Qin, L., 2013. Selective pressure dominates the synonymous codon usage in parvoviridae. Virus Genes 46, 10–19.
Shi, S.-L., Jiang, Y.-R., Yang, R.-S., Wang, Y., Qin, L., 2016. Codon usage in Alphabaculovirus and Betabaculovirus hosted by the same insect species is weak, selection dominated and exhibits no more similar patterns than expected. Infect. Genet. Evol. 44, 412–417.
Siddell, S.G., Ziebuhr, J., Snijder, E.J., 2005. Coronaviruses, Toroviruses, and Arteriviruses. Topley and Wilson's Microbiology and Microbial Infections.
Sueoka, N., 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. 85, 2653–2657.
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, 13769–13780.
Supek, F., 2016. The code of silence: widespread associations between synonymous codon biases and gene function. J. Mol. Evol. 82, 65–73.
Suzuki, H., Brown, C.J., Forney, L.J., Top, E.M., 2008. Comparison of correspondence analysis methods for synonymous codon usage in bacteria. Dna Res. 15, 357–365.
Timani, K.A., Ye, L., Zhu, Y., Wu, Z., Gong, Z., 2004. Cloning, sequencing, expression, and purification of SARS-associated coronavirus nucleocapsid protein for serodiagnosis of SARS. J. Clin. Virol. 30, 309–312.
van Hemert, F., van der Kuyl, A.C., Berkhout, B., 2016. Impact of the biased nucleotide composition of viral RNA genomes on RNA structure and codon usage. J. Gen. Virol. 97, 2608–2619.
Wasmoen, T.L., Kadakia, N.P., Unfer, R.C., Fickbohm, B.L., Cook, C.P., Chu, H.J., Acree, W.M., 1995. Protection of cats from infectious peritonitis by vaccination with a recombinant raccoon poxvirus expressing the nucleocapsid gene of feline infectious peritonitis virus. In: Talbot, P.J., Levy, G.A. (Eds.), Corona- and Related Viruses. Springer, Boston, MA, pp. 221–228.
Woese, C.R., 2002. On the evolution of cells. Proc. Natl. Acad. Sci. 99, 8742–8747.
Wu, G., Yan, S., 2005. Reasoning of spike glycoproteins being more vulnerable to mutations among 158 coronavirus proteins from different species. J. Mol. Model. 11, 8–16.
Wu, H.S., Hsieh, Y.C., Su, I.J., Lin, T.H., Chiu, S.C., Hsu, Y.F., Lin, J.H., Wang, M.C., Chen, J.Y., Hsiao, P.W., Chang, G.D., Wang, A.H., Ting, H.W., Chou, C.M., Huang, C.J., 2004a. Early detection of antibodies against various structural proteins of the SARS-associated coronavirus in SARS patients. J. Biomed. Sci. 11, 117–126.
Wu, X.D., Shang, B., Yang, R.F., Hao, Y., Hai, Z., Xu, S., Ji, Y.Y., Ying, L., Di Wu, Y., Lin, G.M., 2004b. The spike protein of severe acute respiratory syndrome (SARS) is cleaved in virus infected Vero-E6 cells. Cell Res. 14, 400.
Yu, M., Stevens, V., Berry, J.D., Crameri, G., McEachern, J., Tu, C., Shi, Z., Liang, G., Weingartl, H., Cardosa, J., Eaton, B.T., Wang, L.F., 2008. Determination and application of immunodominant regions of SARS coronavirus spike and nucleocapsid proteins recognized by sera from different animal species. J. Immunol. Methods 331, 1–12.
Zhou, H., Yan, B., Chen, S., Wang, M., Jia, R., Cheng, A., 2015. Evolutionary characterization of Tembusu virus infection through identification of codon usage patterns. Infect. Genet. Evol. 35, 27–33.
Zhou, T., Gu, W., Ma, J., Sun, X., Lu, Z., 2005. Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. Biosystems 81, 77–86.