



Analysis of codon usage patterns and influencing factors in Nipah virus

Supriyo Chakraborty^{a,*}, Bornali Deb^a, Parvin A. Barbhuiya^a, Arif Uddin^b



^a Department of Biotechnology, Assam University, Silchar 788011, Assam, India

^b Department of Zoology, Moinul Hoque Choudhury Memorial Science College, Algapur, Hailakandi 788150, Assam, India

ARTICLE INFO

Keywords:

Codon usage bias
Nipah virus
Natural selection
Mutation pressure

ABSTRACT

Codon usage bias (CUB) is the unequal usage of synonymous codons of an amino acid in which some codons are used more often than others and is widely used in understanding molecular biology, genetics, and functional regulation of gene expression. Nipah virus (NiV) is an emerging zoonotic paramyxovirus that causes fatal disease in both humans and animals. NiV was first identified during an outbreak of a disease in Malaysia in 1998 and then occurred periodically since 2001 in India, Bangladesh, and the Philippines. We used bioinformatics tools to analyze the codon usage patterns in a genome-wide manner among 11 genomes of NiV as no work was reported yet. The compositional properties revealed that the overall GC and AT contents were 41.96 and 58.04%, respectively *i.e.* Nipah virus genes were AT-rich. Correlation analysis between overall nucleotide composition and its 3rd codon position suggested that both mutation pressure and natural selection might influence the CUB across Nipah genomes. Neutrality plot revealed natural selection might have played a major role while mutation pressure had a minor role in shaping the codon usage bias in NiV genomes.

1. Introduction

Degeneracy or redundancy of the genetic code ensures that multiple codons codify the same amino acid except for two amino acids *i.e.* methionine and tryptophan. The codons that encode the same amino acid are called synonymous codons. Numerous previous studies have shown that the usage of these synonymous codons in mRNA molecules in varying frequencies leads to a phenomenon known as codon usage bias (CUB) (Hasegawa et al., 1979). The evolution of CUB is very complex and a highly debatable subject. Various evolutionary processes explain the origin of synonymous codon usage variation or CUB, among them the two most accepted theories are the neutral theory and the selection-mutation drift balance model theory (Duret and Mouchiroud, 1999; Sharp et al., 1986, 1993; Shields and Sharp, 1987). However, the impact of these evolutionary forces in different species remains undetermined (Akashi, 1997; Hershberg and Petrov, 2008). In addition, various biological factors have been identified to be associated with CUB such as GC composition, gene expression level, gene length, protein structure, tRNA abundance and its types, hydrophobicity and hydrophilicity of the protein (Bains, 1987; Bernardi and Bernardi, 1986; Gouy and Gautier, 1982; Ikemura, 1981; Tao and Dafu, 1998).

The relationships of codon usage between viruses and their hosts are fascinating as it has significance in overall viral existence, its codon adaptation to host, evasion of host's immune system by viral pathogen

and their co-evolution.

CUB can provide significant insights relating to functional regulation of gene expression level, identification of horizontally transferred genes, optimization of protein expression level and adaptation of pathogens to certain specific hosts (Chaney and Clark, 2015; Lithwick and Margalit, 2005; Liu et al., 2012).

Nipah virus (NiV), an emerging zoonotic paramyxovirus, possesses high pathogenicity that causes fatal disease in both animals and humans (Wong et al., 2002). NiV is a single stranded RNA virus which belongs to genus Henipavirus, within the family Paramyxoviridae (Chua et al., 2000). Genome size varies from 18246 to 18252 nucleotides and the number of genes varies from 6 to 9 (<http://www.ncbi.nlm.nih.gov>). NiV was first identified during an outbreak of a disease that took place in Malaysia in 1998 (Lee et al., 1999). Outbreaks of NiV have occurred periodically since 2001 in India, Bangladesh, and the Philippines (Arankalle et al., 2011; Ching et al., 2015; Hossain et al., 2008; Hsu et al., 2004). In the mid of 2018, Nipah outbreak was again reported in Southern parts of India. The natural hosts of the virus are fruit bats of *Pteropodidae* family (Olson et al., 2002). It is transmitted through contact with NiV infected animal causing a prominent risk for epidemic outbreak or through consumption of contaminated foods. Human-to-human transmission has also been observed (Clayton, 2017; Escaffre et al., 2013). In the outbreaks of Malaysia and Singapore, pigs were reported to be the intermediate hosts, whereas in Bangladesh it was the

* Corresponding author.

E-mail address: supriyo_ch_2008@rediffmail.com (S. Chakraborty).

date palm sap contaminated by infected fruit bats (Chua et al., 2001; Clayton et al., 2016). Clinical features are highly variable and may range from asymptomatic infection through acute respiratory syndrome to fatal encephalitis in humans and respiratory diseases in swine (Harcourt et al., 2001).

To date, there are neither therapeutic nor prophylactic treatments against NiV outbreak. Nonetheless, low mortality rate was observed while ribavirin was administered during Malaysian outbreak (Chong et al., 2001), but the same drug was ineffective in preventing the death of NiV infected hamster (Georges-Courbot et al., 2006). Treatment against NiV is limited to only supportive and preventive care.

Analysis of codon usage patterns in different genomes of NiV would improve our understanding on the mechanism of codon distribution and variation in NiV genomes as a part of its molecular biology and on the factors influencing the codon usage patterns. In this paper, we report the CUB and compositional properties of ORFs (open reading frame) in eleven NiV genomes. This study might give insights into pathogen evolution and disease propagation. Besides, the biological relevance of this study might throw light on the therapeutic interventions relating to NiV.

2. Materials and methods

2.1. Sequence data

The complete ORFs of eleven genomes of Nipah virus were retrieved in FASTA format from GenBank database of the National Center for Biotechnology Information (NCBI) accessible from the website <http://www.ncbi.nlm.nih.gov> (Table 1). Each ORF with exact multiple of three nucleotides having appropriate start and stop codon was analyzed in the present study. Nipah is a RNA virus; nucleotide U in RNA genome is replaced by T in NCBI databases and nucleic acid sequences are presented conventionally in 5'-3' orientation following Wisconsin system (where T represents U). Wisconsin system has added advantage of representing the ORF and the mRNA of a gene using the same sequence, here the letter T represents U. So we used NCBI format of ORFs of genes for analysis, thereby representing U by T.

2.2. Nucleotide composition

Overall nucleotide composition (A, C, T and G%) and nucleotide composition at the third codon position of each ORF (A3, C3, T3 and G3%) in each genome were analyzed. The overall GC content, the GC content at the 1st, 2nd and 3rd codon positions were determined for all eleven genomes of NiV. Also, various nucleotide skews namely AT skew (A-T/A+T), GC skew (G-C/G+C), purine skew (A-G/A+G), pyrimidine skew (T-C/T+C), amino skew (A-C/A+C) and keto skew (T-G/T+G) values were also computed for each ORF to understand the dynamics of nucleotide usage in different NiV genomes.

Table 1
Accession number and acronym of 11 genomes of Nipah virus.

Nipah Virus	Acronym	GenBank Acc No.	Genome size (nt)	Cds
USA_NP_112021.1_1	NiV USA	NC_002728.1	18,246	9
Malaysia_UMMC2_AAK50548.1_1	NiV M1	AY029768.1	18,246	8
Malaysia_UMMC1_AAK50540.1_1	NiV M2	AY029767.1	18,246	8
Malaysia_AAF73377.1_1	NiV M3	AF212302.2	18,246	8
Malaysia_CAF25493.1_1	NiV M4	AJ627196.1	18,246	6
Malaysia_NV/MY/99/UM-0128_CAD92359.1_1	NiV M5	AJ564623.1	18,246	6
Malaysia_NV/MY/99/VRI-2794_CAD92347.1_1	NiV M6	AJ564621.1	18,246	6
Bangladesh_NIVBDG2008RAJBARI_AEZ01384.1_1	NiV B1	JN808863.1	18,252	9
Bangladesh_NIVBDG2008MANIKGONJ_AEZ01379.1_1	NiV B2	JN808857.1	18,252	9
Bangladesh_AAY43911.1_1	NiV B3	AY988601.1	18,252	9
India_Ind-Nipah-07-FG_ACT32611.1_1	NiV I	FJ513078.1	18,252	6

2.3. Relative synonymous codon usage

Relative synonymous codon usage (RSCU) is the ratio of observed frequency to the expected frequency of codons within the synonymous codon family if all the codons for the particular amino acid are used randomly in mRNA molecules. RSCU value < 1.0 implies that the codon is used less frequently than expected and vice-versa. RSCU value > 1.6 identifies the over represented codon while RSCU value < 0.6 reveals the underrepresented codon in the ORF (Behura and Severson, 2012; Sharp and Li, 1986).

RSCU is determined mathematically as

$$RSCU_{ij} = \frac{X_{ij}}{\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}}$$

Here, X_{ij} is the frequency of prevalence of the j^{th} codon for i^{th} amino acid (any X_{ij} with a value of zero is arbitrarily allotted a value of 0.5), and n_i is the total number of synonymous codons that encode the i^{th} amino acid.

2.4. Effective number of codons (ENC)

This parameter is broadly used for measuring the degree of CUB. It estimates the magnitude of deviation of codon usage of a gene from the equal usage of synonymous codons. The ENC value ranges from 20 (highest bias) to 61 (lowest or no bias). High ENC value is an indication of low CUB while a low ENC value signifies high CUB of a gene. ENC value less than 35 is regarded as the significant CUB of a gene (Wright, 1990).

ENC is calculated in standard genetic code (Translation Table 1 of NCBI) as:

$$ENC = 2 + \frac{9}{F_2} + \frac{1}{F_3} + \frac{5}{F_4} + \frac{3}{F_6}$$

Here, F_k ($k = 2, 3, 4$ or 6) is the average of the F_k values for k -fold degenerate amino acids. The F value indicates the probability of two randomly chosen codons being identical for an amino acid.

2.5. Correspondence analysis (COA)

Correspondence analysis on relative synonymous codon usage is often used in CUB analysis to elucidate the major trends in variation of data (Greenacre, 1984). The axes of relative inertia are the major trends.

2.6. Parity rule 2

The Parity Rule 2 (PR2) bias is drawn with AT-bias [A/(A+T)] on the y-coordinate and GC-bias [G/(G+C)] on the x-coordinate where the point 0.5 is the centre of the plot, indicating that no bias exists between the complementary nucleotides.

2.7. Neutrality plot

Neutrality plot is the graphical representation of GC12 versus GC3 which measures the mutation-selection equilibrium in shaping the codon usage variation. Each dot represents an independent gene in the plot. The influence of mutation pressure on the CUB of a gene is indicated from the slope of the regression line of GC12 on GC3 approaching 1 (Sueoka, 1988). However, natural selection is assumed to play a significant role in CUB if the points are widely scattered in the plot.

2.8. Mutational responsive index (MRI)

MRI is the measure of the mutational drift in codons (Gouy and Gautier, 1982). It is a component of CUB which is based on compositional properties of the gene. A positive MRI value indicates directional mutation pressure but a negative MRI value indicates that translational selection operates on the gene (Gatherer and McEwan, 1997).

2.9. Translational selection (P2)

P2 measures the codon-anticodon interaction efficiency and detects the translational efficiency of a gene. It is calculated as

$$P2 = (WWC + SSU)/(WWY + SSY)$$

Where W = A or T, S = C or G and Y = C or T

P2 value > 0.5 indicates a bias in favor of translational selection according to Gouy and Gautier (1982).

2.10. Software used in the study

Pearson's correlation analysis was performed to investigate the relationship of ENC (codon usage bias) with overall GC content, codon-position-specific GC contents (GC1, GC2, GC3), nucleotide skews, amino acids; and also the relationships among overall nucleotide compositions and the compositions at third position of codons using SPSS software (SPSS Inc., Chicago, Illinois, USA) (Chakraborty et al., 2017). Correspondence analysis was performed using PAST software (Paleontological statistics software package) (Deb et al., 2018). All indices of codon usage bias were estimated using a program written in PERL computer language by the corresponding author (SC).

3. Results

3.1. Nucleotide composition of Nipah genomes

The distribution and usage of synonymous codons in coding sequences are greatly influenced by nucleotide composition of a genome (Jenkins and Holmes, 2003). We therefore, analyzed nucleotide composition of ORFs of Nipah genomes as shown in Fig. 1. The mean A% was the highest (32.98), followed by T% (25.07), G% (21.92) and C% (20.03) across all genomes. The overall AT and GC contents were 58.04% and 41.96% respectively, indicating NiV genes were AT rich. We estimated GC contents at different codon positions and found mean GC1% (49.07), GC2% (38.72) and GC3% (38.08) indicating, first codon position in genes possessed almost equal AT1% and GC1% as compared to second and third codon positions.

3.2. Codon usage bias of genes in Nipah genomes

To analyze the extent of variations in CUB, the ENC values of ORFs in each Nipah virus genome were estimated (S1). The ENC values of 11 genomes of NiV were shown in Table 2 and the mean value of all genomes was 51.87, which suggested that the CUB of NiV genomes was low (Butt et al., 2014).

3.3. Codon usage pattern in Nipah genomes

Relative synonymous codon usage (RSCU) values of 59 synonymous codons in 11 genomes of Nipah virus were estimated. The patterns of codons usage were very similar in most of the genomes as shown in Fig. 2. Moreover, RSCU values of genes indicated the nucleotides A and T were more preferred to G and C at third codon position. This result along with nucleotide compositional analysis suggests mutational pressure might affect the codon usage pattern in Nipah genomes. In our analysis, RSCU values of most of the codons were in the range 0.6–1.6, indicating stable genetic composition of the genes. However, a few codons were overrepresented (RSCU > 1.6) i.e. ACA, GCA, AGA, GGA, AGG, CCT, GTT and TCA as depicted in Fig. 3. On the other hand, a few codons were underrepresented (RSCU < 0.6) i.e. TCC, TCG, CGC, CGG, CGT, ACG, GCG, CCG, GGC, TGC and CCC as shown in Fig. 4. Comparison of overrepresented and underrepresented codons of 11 genomes of NiV is shown in Fig. 5.

3.4. Correspondence analysis of Nipah genomes

To decipher the rate of codon usage variations in different genomes of Nipah virus, we performed correspondence analysis with RSCU values of codons. The contributions of Axis 1 and Axis 2 were different in different genomes as shown in S2. The AT ended codons were much closer to axes with a clustering tendency than GC ended codons, indicating nucleobases under mutational pressure might have a significant role in shaping CUB of genes.

3.5. Role of mutational pressure on codon usage bias in Nipah virus genomes

To estimate the role of mutational pressure on shaping CUB in 11 genomes, we compared correlation coefficients between overall nucleotide compositions (A%, T%, G%, C%, GC%) and nucleotide compositions at third codon position of the genes (A3%, T3%, G3%, C3%, GC3%) (Table 3). Significant correlation was found between them at $p < 0.01$ or $p < 0.05$, indicating mutational pressure might play an important role in shaping the codon usage bias. This analysis suggests compositional properties under mutational pressure might have affected the codon usage patterns. However, on correlating ENC with GC contents (GC%, GC1%, GC2%, GC3% and GC12%) (Table 4), we found, significant correlation of ENC with GC3% in NiV M4 (0.819*), NiV M5 (0.842*), NiV M6 (0.845*) and NiV I (0.909*), indicating higher impact of mutational pressure on these genomes of Nipah virus as compared to other genomes.

3.6. Parity plot analysis of Nipah virus genomes

The parity rule 2 bias plots are used for predicting the relative magnitude of mutational pressure and natural selection on gene composition. If mutational pressure is the cause of CUB then AT and GC will be proportionally distributed among the degenerate codon groups of a gene (Chakraborty et al., 2017). However, if mutational pressure and natural selection both are acting on the CUB, then AT and GC will not be proportionally distributed in degenerate codon groups (Sueoka, 1995). To investigate the role of mutational pressure and natural selection, the relationships between G and C content and between A and T content in 2-fold, 4-fold and 6-fold degenerate codon families were analyzed with PR2 bias plot (S3). Our results for 2-fold, 4-fold and 6-fold families revealed unequal distribution of GC and AT, suggesting that both the evolutionary forces i.e. mutational pressure and natural selection are responsible for shaping the CUB in NiV genomes.

3.7. Role of natural selection on codon usage patterns in Nipah virus

It has been proposed that if codon usage pattern is influenced by

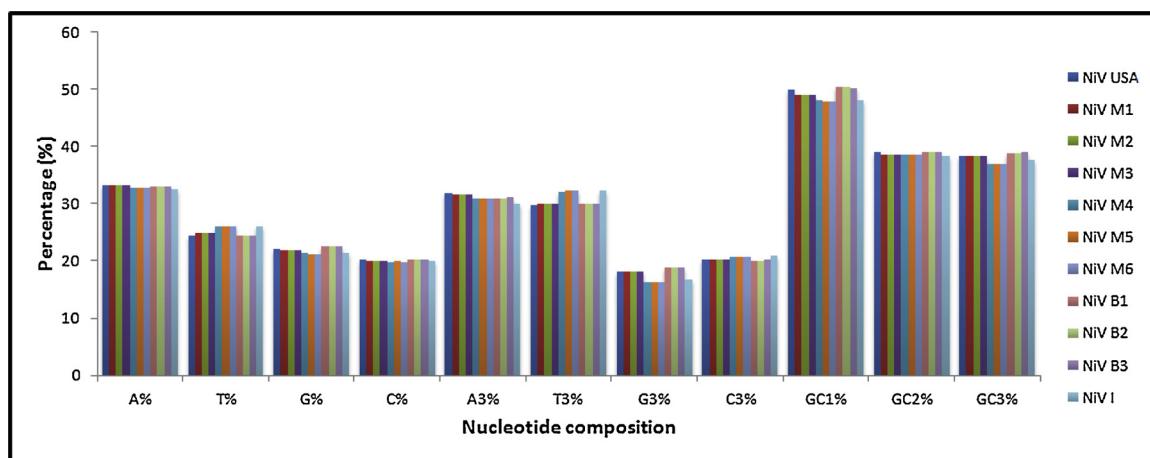


Fig. 1. Overall nucleotide composition and its composition at 3rd codon position in genomes of Nipah virus.

Table 2

Average ENC values of genes in 11 genomes of Nipah virus.

Nipah virus	ENC values
NiV USA	51.87
NiV M1	51.79
NiV M2	51.79
NiV M3	51.79
NiV M4	51.55
NiV M5	51.58
NiV M6	51.57
NiV B1	52.31
NiV B2	52.31
NiV B3	51.96
NiV I	52.05

mutational pressure only, then the frequencies of nucleotides G and C should be equal to that of A and T at third codon position (Zhang et al., 2013a). However in case of Nipah virus genomes, nucleotide composition of GC and AT in the third codon position were unequal (Fig. 1), indicating that natural selection might play a role in shaping CUB.

Further, we drew neutrality plot between GC12 and GC3 (S4) to identify the major determinant of CUB phenomenon between mutational pressure and natural selection and analyzed the regression coefficient (Table 5). The values of regression coefficients were lower than 0.5, indicating selection played a dominant role in shaping the codon usage bias of NiV genomes.

3.8. Relationship between codon usage bias and nucleotide skews

Correlation coefficients were estimated between ENC and various nucleotide skews to unravel their interrelationships in coding sequences. We found significant correlation between ENC and nucleotide skewness (-0.707^*) in NiV USA and (-0.692^*) in NiV B3 at ($p < 0.05$) (S5), thus nucleotide skewness might also affect the codon usage patterns in Nipah virus.

3.9. Mutational responsive index (MRI) and translational selection (P2)

The mean MRI values in 11 genomes of Nipah virus were 0.52 in NiV USA, NiV M1, NiV M2, NiV M3, NiV M4 but 0.51 in NiV M5, NiV M6, NiV B1, NiV B2, NiV B3, NiV I, respectively. Positive MRI value indicates directional mutation pressure (Gatherer and McEwan, 1997). In our study, MRI values in all the Nipah genomes were positive, indicating that directional mutation pressure might have influenced the CUB of NiV genomes. Further, the mean P2 values in 11 genomes of Nipah virus were 0.06 in NiV USA, NiV M1, NiV M2, NiV M3, NiV B1,

NiV B2, NiV B3 and 0.04 in NiV M4, NiV M5, NiV M6, NiV I, respectively. The value of P2 was less than 0.5 (Gouy and Gautier, 1982), indicating that the role of translational selection in CUB was low in NiV genomes. As the MRI and P2 values were almost same in all the Nipah genomes it suggested the effects of directional mutation pressure and translational selection were similar in all genomes.

4. Discussions

CUB arises from unequal distribution of synonymous codons in mature mRNA transcripts. Biased usage of synonymous codons is commonly found in a wide variety of organisms i.e. prokaryotes to eukaryotes (Akashi and Eyre-Walker, 1998; Duret, 2002). For understanding the magnitude and causes of codon bias of genes, the analysis of codon usage and overall nucleotide composition are essential to investigate the evolutionary relatedness, genomic characteristics and the role of mutational pressure and natural selection in genomic composition (Shackelton et al., 2006).

As codon usage bias is highly influenced by nucleotide compositions of a gene, we primarily investigated the nucleotide composition of ORFs in 11 genomes of Nipah virus, and found A% to be the highest, followed by T% > G% > C%; NiV genomes were primarily found to be AT rich. GC contents at three codon positions of genes were found to be GC1% (49.07), GC2% (38.72) and GC3% (38.08). Similar to NiV genomes the percentage of GC1 was the highest compared to GC2 and GC3 in dengue virus, and the variation in GC composition profile was apparently associated to their geographical location (Lara-Ramírez et al., 2014).

For understanding the variations in codon usage we further investigated ENC values of the ORFs of different genomes of Nipah virus and found mean ENC = 51.87 (i.e. ENC much higher than 35) indicating, lower magnitude of CUB. Low CUB might be useful for efficient replication in Nipah virus genomes having different choices for codon usage (Jenkins and Holmes, 2003). Similarly, from the results of codon bias analysis of complete genomic coding sequences in 50 genetically and ecologically diverse human RNA viruses the mean ENC value was found to be 50.9, ranging from 38.9 (Hepatitis A virus) to 58.3 (Eastern equine encephalitis virus), thus depicting lower CUB in most of the viruses, although a few variations were evident (Jenkins and Holmes, 2003). On analysis of synonymous codon usage in SARS Coronavirus, ENC values of genes were found to range from 42.19 to 59.06 with a mean of 48.99, indicating lower codon usage bias.

In order to understand the codon usage patterns, we estimated the RSCU values of 59 synonymous codons and found A/T-ended codons were preferentially used over G/C-ended codons, with a discrete compositional distribution, depicting mutational pressure might play a role in shaping the CUB. In Chikungunya genomes, the codon AGA for

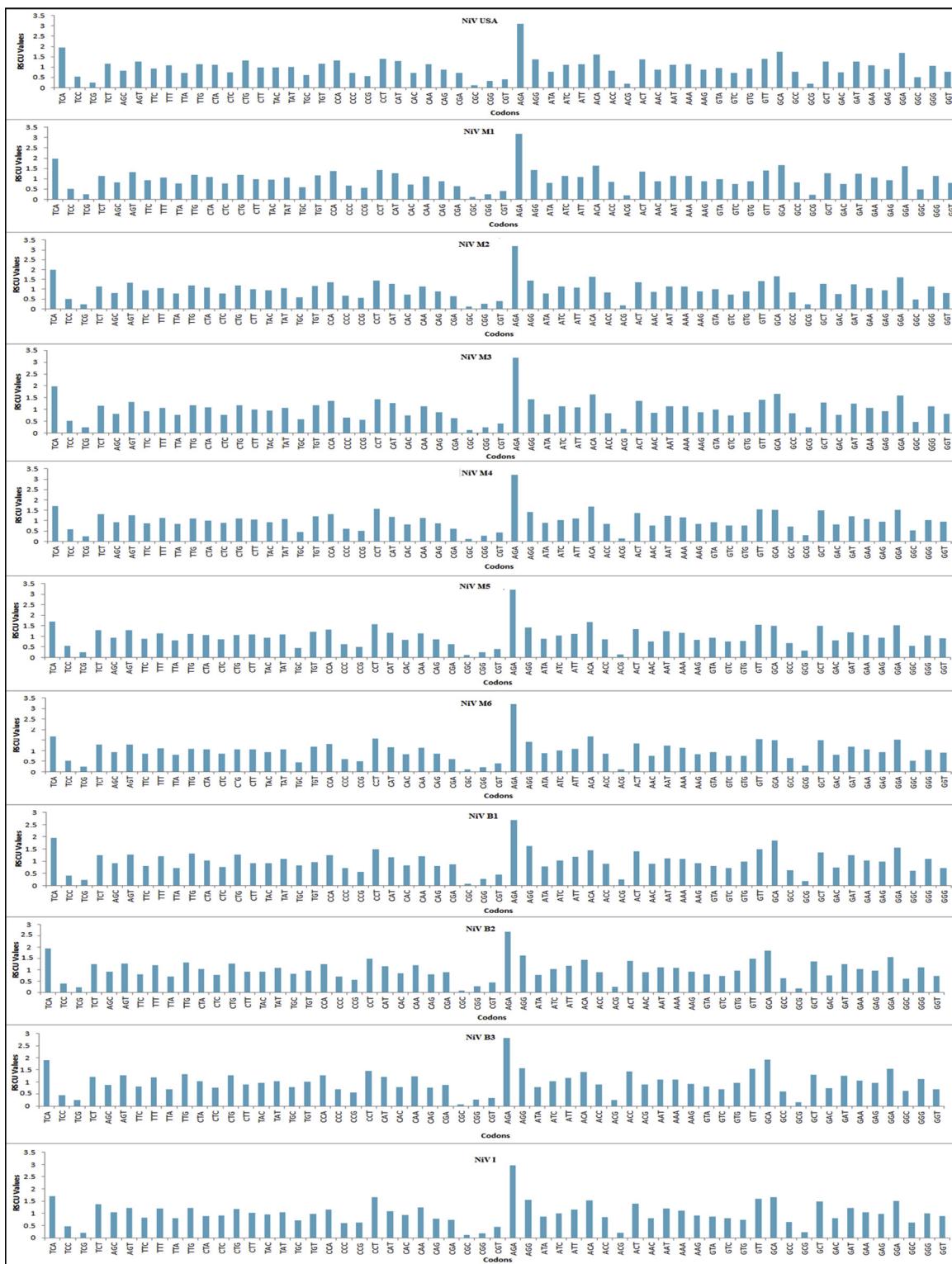


Fig. 2. Comparison of RSCU values of codons in genomes of Nipah virus.

amino acid Arg and CUG for Leu were overrepresented whereas GUU for Val, CGU, CGG for Arg and CUU, CUC for Leu were underrepresented (Butt et al., 2014). Combining the results of nucleotide composition and RSCU analysis in this study, we observed that compositional properties had high impact in selection of preferred codons coupled with the presence of mutational pressure, supporting the result of Butt et al. (2014).

Codon usage variation is multifactorial in nature; therefore for

estimating the rate of variation of codon usage, we used a multivariate statistical technique *i.e.* correspondence analysis. We found mutational pressure might affect the codon usage as GC and AT ended codons were placed close to the axes; however natural selection might also affect the codon bias as some GC and AT ended codons were discretely distributed away from the axes. [Gu et al. \(2004\)](#), from the study of SARS Coronavirus and Nidovirales, reported on correspondence analysis of base compositions and found compositional properties mainly influenced the

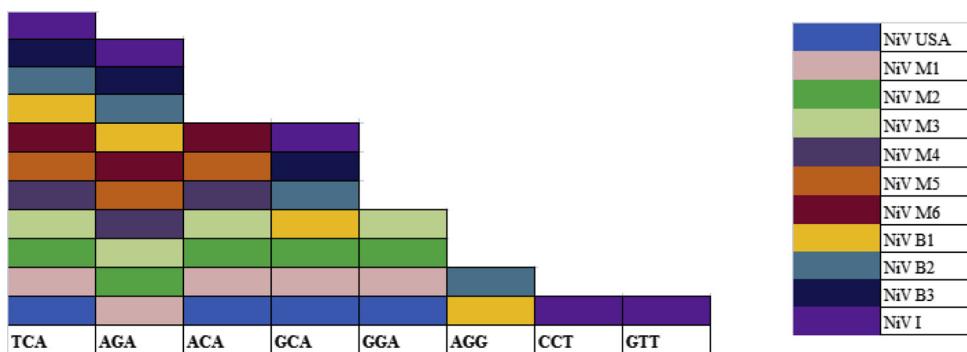


Fig. 3. Overrepresented codons in genomes of Nipah virus.

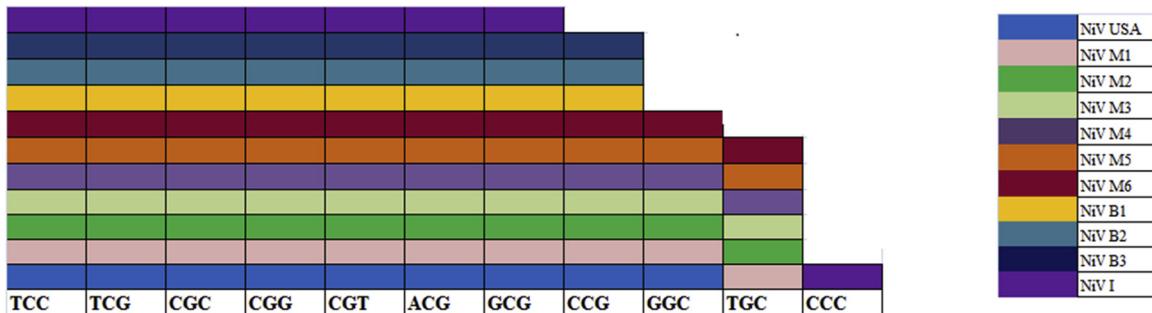


Fig. 4. Underrepresented codons in genomes of Nipah virus.

variations of codon biases in viruses. Moreover, correspondence analysis of Chikungunya viruses revealed that different genotypes were distributed across all planes of axes, indicating the impact of favorable transmission vectors, host range, susceptibility and climatic conditions in shaping codon usage bias and the least influence of mutational pressure (Butt et al., 2014).

All viruses in general and RNA viruses in particular are exposed to high magnitude of mutational pressure (Drake, 1993). Adaption and co-evolution of viruses to their hosts were mostly analyzed by estimating mutation at synonymous and non-synonymous coding sites in specific genes (Yang et al., 2014). For assessing the impact of mutational pressure in codon bias, we correlated overall nucleotide composition (A%, T%, G%, C%, GC%) and nucleotide composition at the third codon position of the genes (A3%, T3%, G3%, C3%, GC3%) using Pearson correlation method and found significant correlation between them. These results depict mutational pressure played a significant role in codon usage of NiV genomes. On correlating ENC with GC3% a significant correlation was obtained at $p < 0.05$ in NiV M4, NiV M5 and NiV M6 and NiV I. Previous studies reported on H5N1 virus and other influenza A viruses, wherein GC composition varied in a similar pattern indicating mutational bias played a major role in codon usage; however the codon usage of NS2 and M2 genes was independent of the GC content (Zhou et al., 2005).

Previous studies reported mutational pressure shaped codon usage in some RNA viruses as mutation rate is much higher in RNA viruses in comparison to DNA viruses (Drake and Holland, 1999). Butt et al. (2014) reported significant correlation among various nucleotide properties at $p < 0.01$ or $p < 0.05$ suggesting, mutational pressure was mostly responsible for nucleotide compositional patterns and for affecting the dynamics of CUB in genes (Butt et al., 2014). To investigate the codon patterns in Zika virus, correlation analysis was performed among nucleotide compositions, and codon compositions in different combinations and significant correlation was reported in them, indicating the role of mutational pressure in codon biases (Butt et al., 2016).

In the present study, the parity rule 2 bias plots analysis was

performed to understand the role of mutational pressure and natural selection on gene composition. We found non-proportional distribution of GC and AT count in 2-fold, 4-fold and 6-fold degenerate codon families, suggesting the role of both mutational pressure and natural selection in codon composition. In Marburg virus, A and T were overrepresented compared to G and C in 4-fold degenerate codon families (Nasrullah et al., 2015). Butt et al. (2016) reported that in Zika virus, nucleobases A and G were more preferred to T and C in 4-fold degenerate families.

If the codon usage is shaped exclusively by mutational pressure, the nucleotide distribution of G and C would be equal to that of A and T at wobble codon position (Zhang et al., 2013b). In our analysis the role of natural selection, apart from mutational pressure, was distinguishable and evident from codon usage patterns in 11 genomes of Nipah virus. To determine the magnitude of mutational pressure over natural selection, neutrality plot was drawn. The results revealed mutational pressure might impact CUB although natural selection was the prime determinant of codon usage in NiV. Similar result was also obtained in Zika virus, with natural selection (96.8%) being the major factor over mutational pressure (3.2%) in shaping the codon bias (Butt et al., 2016). Neutrality plot analysis in Marburg virus revealed mutational pressure (92.6%) was dominant over natural selection (7.4%) for codon usage bias (Nasrullah et al., 2015). Further, Li et al. (2018) reported that in H3N2 canine influenza virus mutational pressure (17.37%) played less significant role than natural selection (82.63%) in codon bias formation.

Nucleotide skewness affects the codon usage bias of genes. Depending on correlation analysis between ENC and nucleotide skews we found significant correlation of codon bias with amino skew in NiV USA and NiV B3, justifying their effects on CUB formation. Berkhouit et al. (2002) reported skew analysis in retroviral genomes that provided information about nucleotide preferences involved in shaping codon usage.

On further analysis of mutational responsive index (MRI) and translational selection (P2) of ORFs, it was observed that mutational pressure might play an effective role over translational selection in 11

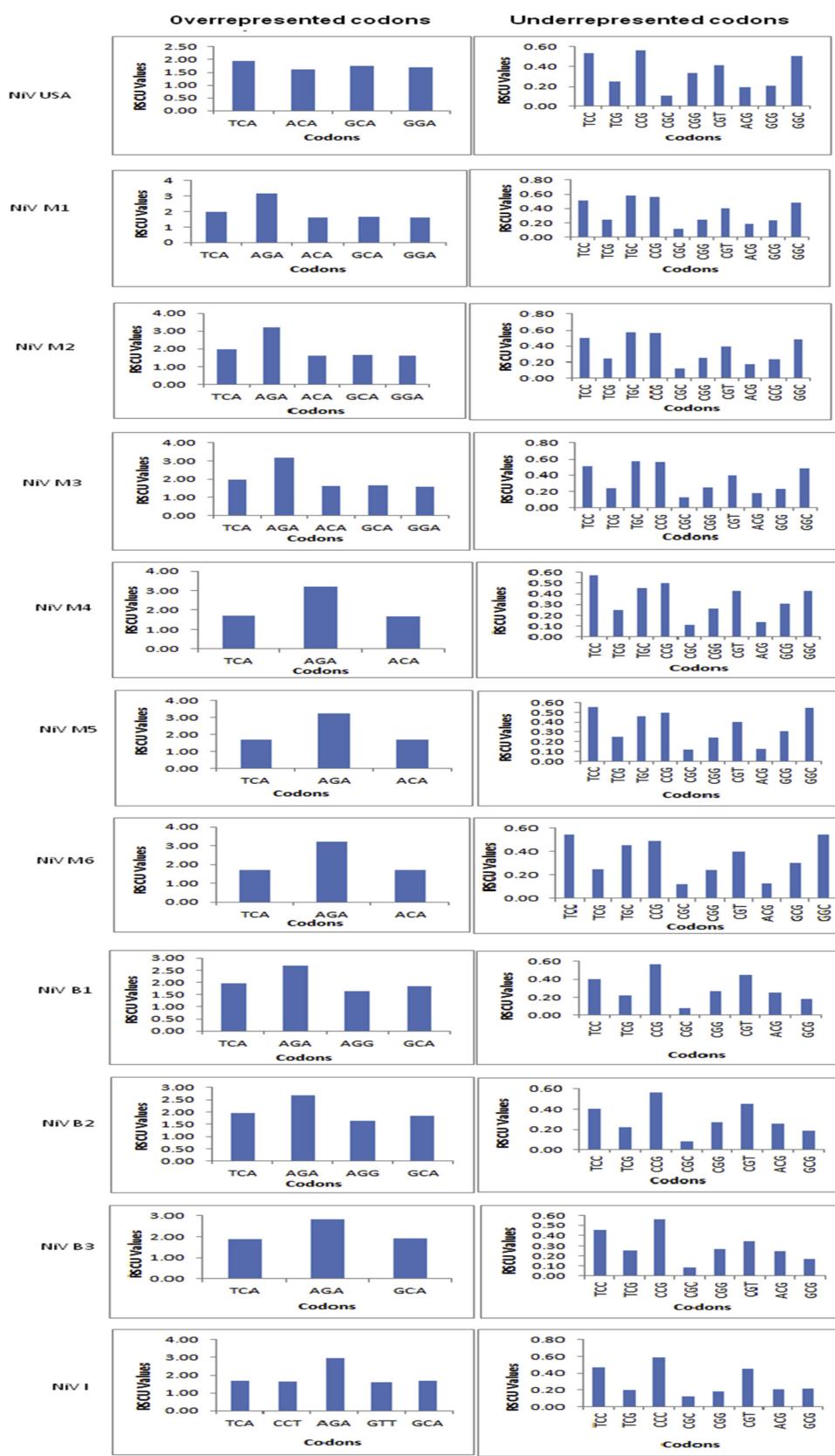


Fig. 5. Comparison of overrepresented and underrepresented codons in genomes of Nipah virus.

Table 3Correlation analysis between overall nucleotide composition and its composition at the 3rd codon position in coding sequences of Nipah virus.

Nipah Virus	Correlation	A3%	T3%	G3%	C3%	GC3%
NiV USA	A%	0.779*	−0.472	0.237	−0.888**	−0.147
	T%	−0.530	0.689*	−0.355	0.221	−0.280
	G%	0.107	−0.554	0.398	0.132	0.484
	C%	0.237	−0.258	−0.081	0.355	0.077
	GC%	0.147	−0.472	0.25	0.223	0.368
	A%	0.777*	−0.463	0.24	−0.884**	−0.139
NiV M1	T%	−0.490	0.715*	−0.394	0.164	−0.346
	G%	0.033	−0.556	0.425	0.207	0.543
	C%	0.163	−0.24	−0.085	0.466	0.117
	GC%	0.068	−0.471	0.271	0.315	0.428
	A%	0.777*	−0.465	0.240	−0.886**	−0.139
	T%	−0.490	0.714*	−0.394	0.172	−0.346
NiV M2	G%	0.033	−0.554	0.425	0.200	0.543
	C%	0.163	−0.237	−0.085	0.461	0.117
	GC%	0.068	−0.468	0.271	0.308	0.428
	A%	0.777*	−0.465	0.240	−0.886**	−0.139
	T%	−0.490	0.714*	−0.394	0.172	−0.346
	G%	0.033	−0.554	0.425	0.200	0.543
NiV M3	C%	0.163	−0.237	−0.085	0.461	0.117
	GC%	0.068	−0.468	0.271	0.308	0.428
	A%	0.777*	−0.465	0.240	−0.886**	−0.139
	T%	−0.490	0.714*	−0.394	0.172	−0.346
	G%	0.033	−0.554	0.425	0.200	0.543
	C%	0.163	−0.237	−0.085	0.461	0.117
NiV M4	GC%	0.068	−0.468	0.271	0.308	0.428
	A%	0.819*	−0.107	−0.73	−0.894*	−0.924**
	T%	−0.310	0.856*	−0.206	−0.037	−0.123
	G%	−0.256	−0.561	0.659	0.498	0.634
	C%	−0.078	−0.745	0.366	0.538	0.526
	GC%	−0.211	−0.639	0.586	0.535	0.625
NiV M5	A%	0.795	−0.189	−0.652	−0.873*	−0.895*
	T%	−0.355	0.855*	−0.132	−0.014	−0.076
	G%	−0.211	−0.522	0.585	0.464	0.591
	C%	0.004	−0.749	0.210	0.526	0.457
	GC%	−0.158	−0.610	0.489	0.511	0.578
	A%	0.795	−0.186	−0.652	−0.872*	−0.892*
NiV M6	T%	−0.351	0.857*	−0.135	−0.023	−0.088
	G%	−0.211	−0.527	0.585	0.469	0.596
	C%	0	−0.748	0.215	0.53	0.467
	GC%	−0.158	−0.616	0.489	0.517	0.585
	A%	0.807**	−0.334	0.092	−0.753*	−0.232
	T%	−0.532	0.595	−0.457	0.615	−0.184
NiV B1	G%	0.128	−0.478	0.487	−0.305	0.344
	C%	0.12	−0.219	0.123	−0.006	0.12
	GC%	0.119	−0.431	0.418	−0.232	0.309
	A%	0.807 **	−0.334	0.092	−0.753*	−0.232
	T%	−0.532	0.595	−0.457	0.615	−0.184
	G%	0.128	−0.478	0.487	−0.305	0.344
NiV B2	C%	0.12	−0.219	0.123	−0.006	0.12
	GC%	0.119	−0.431	0.418	−0.232	0.309
	A%	0.810**	−0.377	0.107	−0.753*	−0.221
	T%	−0.522	0.618	−0.443	0.601	−0.196
	G%	0.108	−0.501	0.494	−0.314	0.377
	C%	0.133	−0.155	0.012	0.077	0.045
NiV B3	GC%	0.106	−0.439	0.401	−0.217	0.322
	A%	0.820*	−0.457	−0.713	−0.74	−0.859*
	T%	−0.428	0.75	−0.244	0.334	0.165
	G%	−0.168	−0.252	0.733	0.104	0.367
	C%	−0.085	−0.514	0.28	0.36	0.397
	GC%	−0.154	−0.34	0.646	0.182	0.396

**, * correlation significant at $p < 0.01$, $p < 0.05$ (2-tailed).

genomes of Nipah virus.

5. Conclusion

Nipah virus (NiV) outbreak from time to time frightens people due to its fatal characteristics. It claims life almost every year. Till date no effective drug as a prophylactic or therapeutic measure against NiV is yet available, apart from only supportive and preventive care.

The present study on 11 genomes of NiV revealed that the NiV genes are AT-rich (58.04%) and the overall CUB of genes was low ($ENC > 51$), suggesting that almost all synonymous codons are used in the ORFs. This suggests the existence of genetic variability in codon usage in NiV genomes. Further, neutrality plot revealed that natural selection might have played a prominent role than mutation pressure in

shaping the CUB of genes during evolution of Nipah virus. Codon bias analysis could be harnessed in designing peptide vaccine against disease causing viruses based on the expression potential of surface proteins, for example by identifying efficient epitopes of surface proteins for boosting up cellular immunity.

Conflict of interests

The authors declared no conflict of interests in the manuscript.

Acknowledgment

We are thankful to Assam University, Silchar, Assam India for providing the necessary facilities to carry out the research work.

Table 4
Correlation coefficient between ENC and various GC contents in Nipah virus.

Nipah virus	ENC and GC%	ENC and GC1%	ENC and GC2%	ENC and GC3%	ENC and GC12%
NiV USA	0.616	0.616	0.369	0.170	0.558
NiV M1	0.616	0.636	0.349	0.183	0.563
NiV M2	0.616	0.636	0.352	0.183	0.565
NiV M3	0.616	0.636	0.349	0.183	0.563
NiV M4	0.630	0.526	0.112	0.819*	0.397
NiV M5	0.597	0.478	0.080	0.842*	0.343
NiV M6	0.599	0.480	0.083	0.845*	0.348
NiV B1	0.342	0.466	0.442	−0.243	0.486
NiV B2	0.342	0.466	0.442	−0.243	0.486
NiV B3	0.275	0.335	0.324	−0.129	0.354
NiV I	0.403	0.159	0.029	0.909*	0.110

**,* correlation significant at $p < 0.01$, $p < 0.05$ (2-tailed).

Table 5
Regression coefficients of GC12 on GC3 content in genomes of Nipah virus.

Nipah virus	Regression coefficients of GC12 on GC3
NiV USA	0.1363
NiV M1	−0.1052
NiV M2	−0.1023
NiV M3	−0.1052
NiV M4	0.3323
NiV M5	0.2935
NiV M6	0.3012
NiV B1	−0.2339
NiV B2	−0.2339
NiV B3	−0.216
NiV I	0.0991

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.virusres.2019.01.011>.

References

Akashi, H., 1997. Codon bias evolution in *Drosophila*. Population genetics of mutation-selection drift. *Genet.* 205, 269–278.

Akashi, H., Eyre-Walker, A., 1998. Translational selection and molecular evolution. *Curr. Opin. Genet. Dev.* 8, 688–693.

Arankalle, V.A., Bandyopadhyay, B.T., Ramdasi, A.Y., Jadi, R., Patil, D.R., Rahman, M., Majumdar, M., Banerjee, P.S., Hati, A.K., Goswami, R.P., 2011. Genomic characterization of Nipah virus, West Bengal, India. *Emerg. Infect. Dis.* 17, 907.

Bains, W., 1987. Codon distribution in vertebrate genes may be used to predict gene length. *J. Mol. Biol.* 197, 379–388.

Behura, S.K., Severon, D.W., 2012. Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. *PLoS One* 7, e43111.

Berkhout, B., Grigoriev, A., Bakker, M., Lukashov, V.V., 2002. Codon and amino acid usage in retroviral genomes is consistent with virus-specific nucleotide pressure. *AIDS Res. Hum. Retroviruses* 18, 133–141.

Bernardi, G., Bernardi, G., 1986. Compositional constraints and genome evolution. *J. Mol. Evol.* 24, 1–11.

Butt, A.M., Nasrullah, I., Tong, Y., 2014. Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. *PLoS One* 9, e90905.

Butt, A.M., Nasrullah, I., Qamar, R., Tong, Y., 2016. Evolution of codon usage in Zika virus genomes is host and vector specific. *Emerg. Microbes Infect.* 5, e107.

Chakraborty, S., Nag, D., Mazumder, T.H., Uddin, A., 2017. Codon usage pattern and prediction of gene expression level in *Bungarus* species. *Gene* 604, 48–60.

Chaney, J.L., Clark, P.L., 2015. Roles for synonymous codon usage in protein biogenesis. *Annu. Rev. Biophys.* 44, 143–166.

Ching, P.K.G., de Los Reyes, V.C., Sucaldito, M.N., Tayag, E., Columna-Vingno, A.B., Malbas Jr, F.F., Bolo Jr., G.C., Sejvar, J.J., Eagles, D., Playford, G., 2015. Outbreak of henipavirus infection, Philippines, 2014. *Emerg. Infect. Dis.* 21, 328.

Chong, H.T., Kamarulzaman, A., Tan, C.T., Goh, K.J., Thayaparan, T., Kunjapan, S.R., Chew, N.K., Chua, K.B., Lam, S.K., 2001. Treatment of acute Nipah encephalitis with ribavirin. *Ann. Neurol.* 49, 810–813.

Chua, K., Bellini, W., Rota, P., Harcourt, B., Tamin, A., Lam, S., Ksiazek, T., Rollin, P., Zaki, S., Shieh, W.-J., 2000. Nipah virus: a recently emergent deadly paramyxovirus. *Science* 288, 1432–1435.

Chua, K., Lam, S., Goh, K., Hooi, P., Ksiazek, T., Kamarulzaman, A., Olson, J., Tan, C., 2001. The presence of Nipah virus in respiratory secretions and urine of patients during an outbreak of Nipah virus encephalitis in Malaysia. *J. Infect.* 42, 40–43.

Clayton, B.A., 2017. Nipah virus: transmission of a zoonotic paramyxovirus. *Curr. Opin. Virol.* 22, 97–104.

Clayton, B.A., Middleton, D., Arkinstall, R., Frazer, L., Wang, L.-F., Marsh, G.A., 2016. The nature of exposure drives transmission of Nipah viruses from Malaysia and Bangladesh in ferrets. *PLoS Negl. Trop. Dis.* 10, e0004775.

Deb, B., Uddin, A., Mazumder, G.A., Chakraborty, S., 2018. Analysis of codon usage pattern of mitochondrial protein-coding genes in different hookworms. *Mol. Biochem. Parasit.* 219, 24–32.

Drake, J.W., 1993. Rates of spontaneous mutation among RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* 90, 4171–4175.

Drake, J.W., Holland, J.J., 1999. Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13910–13913.

Duret, L., 2002. Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* 12, 640–649.

Duret, L., Mouchiroud, D., 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4482–4487.

Escaffre, O., Borisevich, V., Rockx, B., 2013. Pathogenesis of Hendra and Nipah virus infection in humans. *J. Infect. Dev.* 7, 308–311.

Gatherer, D., McEwan, N.R., 1997. Small regions of preferential codon usage and their effect on overall codon bias—the case of the *plp* gene. *IUBMB Life* 43, 107–114.

Georges-Courbot, M., Contamin, H., Faure, C., Loth, P., Baize, S., Leyssen, P., Neyts, J., Deubel, V., 2006. Poly (I)-poly (C12U) but not ribavirin prevents death in a hamster model of Nipah virus infection. *Antimicrob. Agents Chemother.* 50, 1768–1772.

Gouy, M., Gautier, C., 1982. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* 10, 7055–7074.

Greenacre, M.J., 1984. Theory and Applications of Correspondence Analysis.

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.

Harcourt, B.H., Tamin, A., Halpin, K., Ksiazek, T.G., Rollin, P.E., Bellini, W.J., Rota, P.A., 2001. Molecular characterization of the polymerase gene and genomic termini of Nipah virus. *Virology* 287, 192–201.

Hasegawa, M., Yasunaga, T., Miyata, T., 1979. Secondary structure of MS2 phage RNA and bias in code word usage. *Nucleic Acids Res.* 7, 2073–2079.

Hershberg, R., Petrov, D.A., 2008. Selection on codon bias. *Annu. Rev. Genet.* 42, 287–299.

Hossain, M.J., Gurley, E.S., Montgomery, J.M., Bell, M., Carroll, D.S., Hsu, V.P., Formenty, P., Croisier, A., Bertherat, E., Faiz, M., 2008. Clinical presentation of Nipah virus infection in Bangladesh. *Clin. Infect. Dis.* 46, 977–984.

Hsu, V.P., Hossain, M.J., Parashar, U.D., Ali, M.M., Ksiazek, T.G., Kuzmin, I., Niezgoda, M., Rupprecht, C., Bresee, J., Breiman, R.F., 2004. Nipah virus encephalitis reemergence, Bangladesh. *Emerg. Infect. Dis.* 10, 2082.

Ikemura, T., 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* 151, 389–409.

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.

Lara-Ramírez, E.E., Salazar, M.I., López-López, M.J., Salas-Benito, J.S., Sánchez-Varela, A., Guo, X., 2014. Large-scale genomic analysis of codon usage in dengue virus and evaluation of its phylogenetic dependence. *BioMed Res. Int.* 2014.

Lee, K.E., Umapathi, T., Tan, C.B., Tjoei-Lian Tjia, H., Chua, T.S., Oh, H.M.L., Fock, K.M., Kurup, A., Das, A., Tan, A.K.Y., 1999. The neurological manifestations of Nipah virus encephalitis, a novel paramyxovirus. *Ann. Neurol.* 46, 428–432.

Li, G., Wang, R., Zhang, C., Wang, S., He, W., Zhang, J., Liu, J., Cai, Y., Zhou, J., Su, S., 2018. Genetic and evolutionary analysis of emerging H3N2 canine influenza virus. *Emerg. Microbes Infect.* 7.

Lithwick, G., Margalit, H., 2005. Relative predicted protein levels of functionally associated proteins are conserved across organisms. *Nucleic Acids Res.* 33, 1051–1057.

Liu, X.-s., Zhang, Y.-g., Fang, Y.-z., Wang, Y.-l., 2012. Patterns and influencing factor of synonymous codon usage in porcine circovirus. *Virol. J.* 9, 68.

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.

Olson, J.G., Rupprecht, C., Rollin, P.E., An, U.S., Niezgoda, M., Clemens, T., Walston, J., Ksiazek, T.G., 2002. Antibodies to Nipah-like virus in bats (*Pteropus lylei*), Cambodia. *Emerg. Infect. Dis.* 8, 987.

Shackelton, L.A., Parrish, C.R., Holmes, E.C., 2006. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J. Mol. Evol.* 62, 551–563.

Sharp, P.M., Li, W.-H., 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* 24, 28–38.

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, 5125–5143.

Sharp, P.M., Stenico, M., Peden, J.F., Lloyd, A.T., 1993. Codon Usage: Mutational Bias, Translational Selection, or Both? Portland Press Limited.

Shields, D.C., Sharp, P.M., 1987. Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. *Nucleic Acids Res.* 15, 8023–8040.

Sueoka, N., 1988. Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.* 85, 2653–2657.

Sueoka, N., 1995. Instrand parity rules of DNA base composition and usage biases of synonymous codons. *J. Mol. Evol.* 40, 318–325.

Tao, X., Dafu, D., 1998. The relationship between synonymous codon usage and protein structure. *FEBS Lett.* 434, 93–96.

Wong, K., Shieh, W., Zaki, S., Tan, C., 2002. Nipah virus infection, an emerging paramyxoviral zoonosis. Springer Seminars in Immunopathology. Springer, pp. 215–228.

Wright, F., 1990. The 'effective number of codons' used in a gene. *Gene* 87, 23–29.

Yang, X., Luo, X., Cai, X., 2014. Analysis of codon usage pattern in *Taenia saginata* based on a transcriptome dataset. *Parasites Vectors* 7, 527.

Zhang, Z., Dai, W., Dai, D., 2013a. Synonymous codon usage in TTSuV2: analysis and comparison with TTSuV1. *PLoS One* 8, e81469.

Zhang, Z., Dai, W., Wang, Y., Lu, C., Fan, H., 2013b. Analysis of synonymous codon usage patterns in torque teno sus virus 1 (TTSuV1). *Arch. Virol.* 158, 145–154.

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.