A detailed analysis of codon usages bias and influencing factors in the nucleocapsid gene of Nipah Virus

Nimmi Chaudhary
Guru Angad Dev Veterinary and Animal Sciences University
Niraj K. Singh (nirajvet57@gmail.com)
Guru Angad Dev Veterinary and Animal Sciences University https://orcid.org/0000-0001-9810-0517

Research Article

Keywords: Nipah Virus, Mutational Pressure, Natural Selection, Codon usage bias

Posted Date: March 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1411194/v1

Abstract

Several outbreaks of Nipah Virus (NiV) have recently been reported in various parts of the world including India. The nucleocapsid (N) protein is the major structural and regulatory (for viral replication cycle) protein of NiV. In the current study, we have conducted a codon usage analysis of N protein encoding gene (N gene) of NiV. The relative synonymous codon usage (RSCU) values, in combination with an ENC value of 50.98, represented low codon usage bias in N gene. The effect of mutational pressure on codon usage bias was confirmed by significant correlations of GC3s, G3s, C3s, A3s, U3s, and ENC values with whole nucleotide contents (GC%, G%, C%, A%, and U%). Correlation study of GC3s, G3s, C3s, A3s, and U3s with axis values of correspondence analysis (CA) also supported the role of mutational pressure. The correlation study of Gravy values with GC3s, G3s, C3s, A3s, and U3s revealed the presence of natural selection in addition to mutational pressure on codon usage bias. Moreover, NiV codon adaptation index (CAI) value higher than their corresponding expected CAI (eCAI) values against human (CAI, 0.726; eCAI, 0.713), pig (CAI, 0.838; eCAI, 0.819), and bat (CAI, 0.763; eCAI, 0.751) also indicated natural selection play role on codon usage bias. Additionally, geographical distribution, and evolutionary processes also influenced the codon usage bias to some extent.

1. Introduction

Nipah virus (NiV) is a highly contagious zoonotic virus that can infect both wild animals and human beings and is listed under the "Terrestrial Animal Health Code" of the World Organization for Animal Health (OIE) (https://www.oie.int/en/disease/nipah-virus/). It is a single-stranded negative-sense RNA virus of the *Paramyxoviridae* family of genus *Henipavirus*. Twenty-three years ago, NiV was reported [1] and between September 1998 and May 1999 (the first outbreak), it was having 40% mortality rate with loss of 105 lives in Malaysia [2]. In India and Bangladesh, NiV outbreaks were reported with a high fatality rate of 70%, in 2001 [3]. This virus also causes catastrophic infections in animals such as pigs, causing huge financial losses in the piggery industry. Several NiV strains have been reported with varied clinical and epidemiological characteristics [4].

The NiV has a single-stranded negative-sense RNA (ssRNA) genome of 18.2-kb size. The genome has six genes that encoded nine proteins including phosphoprotein (P), nucleoprotein (N), fusion protein (F), glycoproteins (G), large polymerase (L), matrix protein (M), W, V, and C protein [5]. The N protein is the most abundant viral protein [6] which interacts with the P protein of the polymerase complex. Relative availability of N protein is determining factor for the activation of genome encapsidation, replicase activity, and regulating viral RNA synthesis. Overexpression of N protein inhibits *in trans* viral transcription while promoting viral genome synthesis [7]. Therefore, N protein played an important regulatory role in virus replication.

Amino acids are the building blocks of proteins, and 20 amino acids are encoded by a set of 61 different codons. Except for Methionine and Tryptophan, most amino acids are coded by more than one codon due to the degeneracy of genetic codes. The use of more than one codon for each amino acid is referred to as synonymous codon usage. Synonymous codon usages are not random, but some codons are preferred over

others. The non-random usage of synonymous codons is termed codon usage bias. Several factors including mutational pressure, natural selection, geographical distribution, evolutionary process, etc are the main driving forces for codon usage bias in many RNA viruses [8, 9]. It has been reported that virus with several ORFs has varied codon usage patterns for different genes [10, 11].

Considering the public health concern of NiV and the importance of N protein in virus replication and assembly, we examined the codon usage bias of N gene and its influencing factors.

2. Materials And Methods

2.1. Nucleotide sequence data Collection

The N gene sequences (1599 base length) of thirty-two NiV isolates were downloaded from the nucleotide database maintained by the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov) in FASTA format. These sequences were used for the analysis of codon usage indexes and phylogenetic analysis. According to the Wisconsin system, the nucleic acid sequences are presented conventionally in the 5'–3' direction and the nucleotide T in the NCBI database is replaced by U in the RNA genome [12]. Table S1 contains the detailed information (GenBank accession number, geographic location, year of sequence submitted to NCBI, etc) of all sequences used in this study (Supplementary Table 1).

2.2. Whole nucleotide and codon third position nucleotide composition analysis

The nucleotide content (G%, C%, A%, and U%) of the N gene coding region was calculated using the CAI Cal (available at http://genomes.urv.es/CAIcal/). Moreover, nucleotide compositions of synonymous codons at the third position (G3s, C3s, A3s, and U3s,) and G + C contents of synonymous codons at the third position (GC3s) were calculated as explained by Peden [13] with the help of CodonW (version 1.4.2) program (http://codonw.sourceforge.net//).

2.3. Relative synonymous codon usage (RSCU) analysis

The ratio between the observed frequency of a given codon of a gene to the frequency of all the expected synonymous codons for a particular amino acid is indicated by the relative synonymous codon usage (RSCU) value. The use of RSCU matrix for the assessment of codon usage bias was first described by Sharp and Li [14]. The RSCU value lesser than 1.0 represents negative codon usage bias while higher than 1.0 indicates positive codon usage for a given synonymous codon [15]. RSCU values of N gene of NiV isolates and bat (*Pteropus alecto*) were calculated by using the following formula in codonW program:

$$RSCU_x = \frac{Frequency of codon_x}{Expected frequency of codon_x if codon usage was uniform}$$

Here, 'x' represents any codon.

2.4. Effective number of codons analyses

In the coding sequence, most of the amino acids (except Methionine and Tryptophan) are encoded by multiple codons which are known as synonymous codons. The effective number of codons (ENC) can be used for estimation of degree of bias in the codon usage [16], which ranges from 20 to 61. If the ENC value is 61, it indicates no codon usage bias. Whereas if the ENC value is 20, it indicates codon usage bias is at extreme and only one codon is being used from each amino acid [9]. The ENC value was calculated using the following formula in CodonW software:

$$ENC = 2 + s + \left(\frac{29}{s^2 + (1+s)^2} \right)$$

The 's' represents the GC3s value [16].

2.5. Codon adaptation index analysis

Codon adaptation index (CAI) can be used for estimation of synonymous codon usage bias in protein coding nucleic acid sequence of a given gene. It represented the comparison between given gene synonymous codon usage and synonymous codon frequency in a reference set [14]. The CAI is a quantitative value that indicates how many times a preferred codon is used among highly expressed genes. It is an indicator of translational efficiency [17]. CAI values range between 0 and 1. A higher value of CAI indicates higher gene expression potential. The CAI value is independent of sequence length and it depends only on the codon frequency [18]. The effect of hosts (human, pig, and bat) on codon usage of NiV was estimated as per the method described by Puigbo et al. [19] for CAI calculation. Codon usage tables for human (*Homo sapien*) and pig (*Sus scrofa domesticus*) were used from previously published data [9] while for bat (*Pteropus alecto*) was prepared by using CDS sequences available on NCBI (https://www.ncbi.nlm.nih.gov/nuccore/). Host codon usage tables were used for calculating CAI of N gene of various NiV isolates. Occasionally, extreme nucleotide/amino acid compositions may yield statistically irrelevant CAI values. Therefore, Puigbo et al. [19] recommended the metric of expected CAI (eCAI) for statistical analysis of CAI analysis and developed a perl script (CAIcal_ECAI_v1.4.pl). The eCAI of N gene was calculated with a 95% confidence interval, as described by Puigbo et al. [19].

2.6. Aromaticity and hydropathicity analysis

Aromaticity (Aromo) and hydropathicity (Gravy) are the factors that indicate the effect of translation or natural selection of the given gene product. The Gravy value represented the average hydropathy value of amino acids in a protein whereas, the Aromo value represented the frequency of aromatic amino acids [20]. The Aromo and Gravy values were calculated by using the following formula [13]:

$$Gravy = \frac{1}{N} + \sum_{i=1}^{N} k_i$$

$$Aromo = \frac{1}{N} + \sum_{i=1}^{N} v_i$$

Here, 'N' is the number of amino acids, k_i is the hydrophobic index of i^{th} amino acid and v_i is either 1(for an aromatic amino acid) or zero.

2.7. Correspondence analysis

The correspondence analysis (CA) is a multivariate analysis method used for the analysis of complex codon usage data. The data of CA is not only represented in form of rows and columns but also helps in identifying major variable trends in the data. To better understand variations, the output of CA can be plotted along various axes [9]. In this study, we used the CodonW program to perform CA on RSCU. Further, CA of the codon usage pattern of N gene from NiVs of different geographical distributions was also performed. With the help of XLSTAT 2015 software, graphs were plotted using the first two principal axes of CA.

2.8. Mutational pressure and natural selection analysis

For the determination of codon usage bias in viruses, mutational pressure and natural selection are two important factors. Correlations of GC3s, G3s, C3s, A3s, U3s, and ENC values with the nucleotide contents and axes value of CA can be used to evaluate the effect of mutational pressure on codon usage. However, the mutational pressure can also be estimated by correlation analysis of %GC and GC3s values [9]. Furthermore, correlations of the Gravy and the Aroma values with GC3s, G3s, C3s, A3s, and U3s can be used to evaluate the effect of natural selection [21, 22]. In this study, XLSTAT 2015 software was used for all correlation analyses.

2.9. Phylogenetic Analysis

Codon usage bias is also influenced by the evolutionary processes of several viruses [23, 24]. MEGAX software was used to study phylogenetic analysis based on the nucleotide sequence. The nucleotide sequences of the N gene of several NiVs were first aligned using the MEGAX program. Further, the aligned sequences were used to construct the phylogenetic tree with the maximum likelihood method with complete deletion parameters. The Robustness in the phylogenetic tree was tested using the Bootstrap method [9].

3. Results

3.1. General features of N gene of NiV

In the coding sequence of the N gene of NiV isolates, the percentage of G, C, A, and U nucleotides were calculated. The G%, C%, A%, and U%, were 25.10 ± 0.137 (mean ± SD), 20.31 ± 0.165, 31.32 ± 0.210, and 23.25 ± 0.256, respectively. Moreover, GC3s, G3s, C3s, A3s, and U3s were 37.10 ± 0.268 (mean ± SD), 20.92 3 ± 0.723, 25.037 ± 0.611, 39.073 ± 0.961, and 35.887 ± 1.059, respectively (Supplementary Table 1).

Further, ENC values of the N gene of NiV isolates were calculated (Supplementary Table 1) to estimate the degree of codon usage bias in the NiV. The ENC value was 50.98 ± 0.367 (mean ± SD) indicated low codon usage bias in NiV.

3.2. RSCU analysis

To analyze the codon usage bias and effect of the hosts (Human, Pig, and Bat) on the N gene codon usage bias, the RSCU values for each synonymous codon in N gene of NiV isolates and its hosts were calculated and compared (Table 1). Out of 18 amino acids (coded by more than one codon), preferred codons for 2

amino acids [Ile (AUC) and Arg (AGA)] were similar between NiV and humans. Preferred codons for 5 amino acids [Ile (AUC), Pro (CCA), Glu (GAA), Arg (AGA), Gly (GGA)] were similar between NiV and pig, whereas preferred codons for only 1 amino acid [Ile (AUC)] was found to be similar between NiV and bat (Table 1). These fewer common preferred codons between NiV and its hosts indicated codon usage bias in N.

Table 1: The synonymous codon usage pattern of N gene in NiVs and its hosts AA: amino acid

AA	Codons	RSCU				AA	Codons	RSCU			
		NiV	Human	Pig	Bat			NiV	Human	Pig	Bat
Phe	UUU	**1.14**	0.94	0.97	0.95	**Glu**	GAA	**1.18**	0.86	**1.03**	0.90
	UUC	0.86	**1.06**	**1.03**	**1.05**		GAG	0.82	**1.14**	0.97	**1.10**
Leu	UUA	0.47	0.47	0.59	0.78	**Ala**	GCU	**1.77**	1.05	1.10	1.09
	UUG	0.95	0.78	0.91	1.22		GCC	0.63	**1.59**	**1.30**	**1.57**
	CUU	0.99	0.80	1.00	0.71		GCA	1.47	0.92	1.11	0.94
	CUC	**1.82**	1.15	1.16	0.96		GCG	0.14	0.44	0.49	0.40
	CUA	1.14	0.43	0.56	0.39						
	CUG	0.63	**2.36**	**1.79**	**1.93**	**Tyr**	UAU	**1.36**	0.90	0.98	0.92
							UAC	0.64	**1.10**	**1.02**	**1.08**
Met	AUG	1.00	1.00	1.00	1.00						
						Cys	UGU	0.00	0.93	0.90	0.96
Val	GUU	**1.43**	0.74	0.88	0.75		UGC	0.00	**1.07**	**1.10**	**1.04**
	GUC	0.95	0.94	1.00	0.95						
	GUA	0.54	0.48	0.60	0.52	**Trp**	UGG	1.00	1.00	1.00	1.00
	GUG	1.08	**1.84**	**1.52**	**1.77**						
						His	CAU	1.00	0.85	0.95	0.86
Ser	UCU	1.04	1.12	1.03	1.27		CAC	1.00	**1.15**	**1.05**	**1.14**
	UCC	0.23	1.28	1.16	**1.34**						
	UCA	1.66	0.91	1.16	1.02	**Gln**	CAA	**1.16**	0.54	0.79	0.55
	UCG	0.06	0.33	0.40	0.37		CAG	0.84	**1.46**	**1.21**	**1.46**
	AGU	**1.93**	0.91	0.90	0.82						
	AGC	1.07	**1.44**	**1.35**	1.18	**Asn**	AAU	**1.22**	0.96	0.97	0.98
							AAC	0.78	**1.04**	**1.03**	**1.02**
Pro	CCU	1.16	1.14	1.10	1.18						
	CCC	0.51	**1.29**	1.10	**1.24**	**Arg**	CGU	0.32	0.48	0.43	1.03
	CCA	**1.49**	1.10	**1.23**	1.13		CGC	0.00	1.10	0.65	0.97
	CCG	0.85	0.47	0.57	0.45		CGA	0.29	0.65	0.58	0.59
							CGG	0.27	1.22	0.77	1.18

Thr	ACU	**1.80**	1.01	0.97	1.04		AGA	**3.50**	**1.29**	**1.99**	0.82
	ACC	0.84	**1.39**	1.22	**1.33**		AGG	1.63	1.27	1.58	**1.40**
	ACA	1.35	1.15	**1.29**	1.16						
	ACG	0.00	0.46	0.52	0.48	**Gly**	GGU	0.61	0.64	0.63	0.67
							GGC	0.86	**1.35**	1.10	**1.31**
Asp	GAU	**1.03**	0.94	0.97	0.95		GGA	**1.73**	1.01	**1.35**	1.04
	GAC	0.97	**1.06**	**1.03**	**1.05**		GGG	0.80	1.00	0.93	0.98
Lys	AAA	**1.08**	0.88	0.99			AUU	0.86	1.10	1.09	1.11
	AAG	0.92	**1.12**	**1.01**	**1.09**		AUC	**1.29**	**1.38**	**1.20**	**1.34**
						Ile	AUA	0.86	0.52	0.71	0.55

AA: amino acid

The RSCU values in bold letter are the preferentially used codons.

Further, to check the codon usage bias in the N gene of NiV isolates, the correspondence analysis (CA) analysis based on RSCU (CA-RSCU) was performed. The CA-RSCU indicated that the first, second, and third axis contributed for 86.05%, 4.61%, and 2.48% of total variations, respectively. Therefore, the first axis mostly explained the presence of codon usage bias. A graph was plotted using first and second axes values to understand the distribution of synonymous codons. Most preferred and moderately preferred codons in various synonymous groups were found to be closer to the intersection of axis 1 and axis 2 while the least preferred codons in respective synonymous groups were found away from the intersection (Fig. 1).

3.3. Effect of mutational pressure on N gene codon usage bias

To analyze the factors affecting codon usage bias in the N gene, a graph using ENC and GC3s values was plotted. In the graph, a single cluster was observed for all NiVs indicating less variation of ENC values among various isolates (Fig. 2). However, all NiV isolates were lying slightly below the expected curve. This suggested that the codon usage bias of the NiV might be due to a combination of mutational pressure and natural selection.

For evaluation of the degree of codon usage bias influenced by mutational pressure, the correlation was performed among nucleotide composition at the third position of codons, ENC values, and whole nucleotide compositions. Significantly high correlations among third position nucleotide composition of codons, ENC values, and whole nucleotide compositions (excluding poor correlations of whole nucleotide compositions with GC3s values and %G with ENC values) were observed (Table 2), while a weak correlation ($r = 0.402$, $p <$

0.02) between %GC values GC3s was observed (Fig. 3). These results indicated that in addition to mutational pressure, other factors also influence NiV codon usage bias.

Table 2
The correlation (Spearman) between nucleotide compositions (A%, C%, U%, G%) with GC3s, G3s, C3s, A3s, U3s, ENC values, the first axis values, the second axis values, and the Gravy values of N gene of NiV isolates

Variables	Gravy	Axis1	Axis2	U3s	C3s	A3s	G3s	ENC	GC3s
%A	**0.796***	**-0.855***	-0.172**	**-0.917***	**0.881***	**0.977***	**-0.943***	**-0.653***	-0.251**
%C	**0.760***	**-0.790***	-0.122**	**-0.946***	**0.978***	**0.875***	**-0.897***	**-0.507#**	0.082**
%U	**-0.766***	**0.869***	0.123**	**0.988***	**-0.908***	**-0.895***	**0.881***	**0.583#**	-0.031**
%G	**-0.616***	**0.618***	0.323**	**0.727***	**-0.835***	**-0.821***	**0.810***	0.275**	0.101**
*p < 0.0001.									
**p > 0.05.									
#p 0.05 < 0.0001									

3.4. Effect of natural selection on N gene codon usage bias

To analyze the effect of natural selection on the N gene of NiV codon usage bias the correlation analysis between nucleotide composition at the third position of all codons, and ENC values with Aroma values and Gravy values was performed. Aroma values do not have any correlation (due to the absence of variation of aroma value in N gene of various NiV isolates) with the GC3s, G3s, C3s, A3s, U3s, and ENC values. But GC3s, G3s, C3s, A3s, U3s, and ENC values have significantly correlated with the Gravy values (Table 3). These results indicated that in addition to mutational pressure, natural selection has also influenced the codon usage bias of the N gene.

Table 3
The correlation (Spearman) between Gravy values with A3s, U3s, G3s, C3s, GC3s, and ENC values of N gene of NiV isolates.

Variables	U3s	C3s	A3s	G3s	ENC	GC3s
Gravy	**-0.802***	**0.806***	**0.798***	**-0.776***	**-0.564***	-0.125**
*p < 0.0001.						
**p > 0.05.						

Subsequently, the relative adaptiveness of NiV codon usage to its hosts was measured by using the CAI metric. The CAI values of NiV were found to be 0.726 ± 0.003 (mean ± SD), 0.838 ± 0.003, and 0.763 ± 0.004 when compared with human (CAI^H), pig (CAI^P), and bat (CAI^B), respectively (Supplementary Table 1). To lessen the effect of extreme G + C and/or amino acid compositions and to overcome the effects of compositional, Puigbo et al. [19] suggested the use of the eCAI algorithm. Codon usage bias influenced by natural selection was also confirmed by higher CAI values of all NiV isolates than their corresponding eCAI values against human ($eCAI^H$, 0.713), pig ($eCAI^P$, 0.819), and bat ($eCAI^B$, 0.751).

3.5. Effect of geographical distribution and evolutionary process on N gene codon usage bias

Based on the geographical distribution and time of N gene sequence of various isolates reported to NCBI, NiVs were grouped into three different bunches (Fig. 4B). In the first bunch, all Malaysian isolates (sequence reported between 1999 and 2004) were found. In the second bunch, NiV isolates from Cambodia, and Thailand (sequence reported between 2005 and 2013) while in the third bunch, isolates from India and Bangladesh (sequence reported between 2005 and 2011) were reported (Supplementary Table 1). Further, we evaluated the codon usage in NiV isolates from different geographical locations by CA. During CA, the first and second axis contributed 86.05%, and 4.61% of the total variation, respectively. The graph of the first and second axis of CA showed that all NiV isolates were organized into two distinct clusters (Fig. 4C). All isolates of the Malaysia, and Cambodia were found in cluster-A, while all isolates from India and Bangladesh were found in cluster-B. NiV isolates from Thailand were distributed in both cluster-A and Cluster-B. Subsequently, phylogenetic analysis using N gene of various NiV isolates was carried out. In the phylogenetic tree, all NiV isolates were organized into two separate clades (Fig. 4A) having similarities to the clustering pattern observed in CA analysis (Fig. 4C). The similarity in geographical distributions, evolutionary tree, and CA results of various isolates indicated the influence of geographical distribution and evolutionary processes on codon usage bias.

4. Discussion

The NiV is an emerging bat-borne pathogen that causes severe respiratory and neurological disease with high mortality. It can spread in the population through infected people or infected animals. Based on epidemiological distribution, different strains of the virus with differing clinical features have been reported [1]. Information of factors influencing codon usage bias and its intensity are important to know detail about viral evolution and its transmission. Previously, Khandia et al. [3]and Chakraborty et al. [12] studied the NiV codon usage pattern and its influencing factors. In both studies, RSCU values (a major indicator of codon usage bias) and codon usage patterns for the complete genome were evaluated. But, it has been reported that virus with different open reading frames (ORFs) has varied codon usage patterns for different genes [10, 11].

The NiV genome (single-stranded negative-sense RNA; 18.2-kb size) has six genes/ORFs which encoded nine different proteins (N, P, F, G, W, V, C, M, and L proteins) [5]. As, the N gene encodes for viral N protein, which is the most abundant protein among all structural proteins of NiV [6], and relative abundance of N

protein is a major controlling factor for genome encapsidation, replicase activity, and regulating viral RNA synthesis [7], the current study was focused to understand the codon usage bias of the N gene of NiV using multiple systemic analytical methodologies. We calculated and compared RSCU values for each synonymous codon of the N gene of various NiV isolates and its hosts (Human, Pig, and Bat). Comparison of the preferred codon of each amino acid of viral N gene and its host indicated 2 preferred codons [Ile (AUC) and Arg (AGA)] were common between virus and humans whereas 5 preferred codons [Ile (AUC), Pro (CCA), Glu (GAA), Arg (AGA), Gly (GGA)] were common between virus and pig, and only 1 preferred codon [Ile (AUC)] was common between virus and bat. RSCU comparison of viral N gene and its host indicates the presence of codon usage bias in NiV.

The ENC is a simple indicator of codon bias. Earlier, ENC values of various RNA viruses have been determined like Japanese encephalitis virus (mean ENC = 55.30) [9] Zika virus (mean ENC = 52.72) [8], and chikungunya viruses (mean ENC = 55.56) [21]. The higher (more than 45) ENC value is an indicator of week codon usage bias [8, 9]. In the current study, the mean ENC value for the N gene of NiV isolates was 50.98 indicating low codon usage bias in NiV. The virus having low codon usage bias can use multiple codons for each amino acid which allows viral replication more efficiently in the host cell [22].

In several RNA viruses, mutational pressure and natural selection are two key forces that determine codon usage bias [21]. If mutational pressure is the only factor determining codon usage bias, during the ENC versus GC3 analysis, all data points reflecting ENC values should lie on the expected curve [16]. In this study, the data points were found below the predicted curve. This indicated that other factors also influence codon usage bias of the N gene of NiV in addition to the mutational pressure. The effect of mutational pressure on codon usage bias was supported by substantial correlations between total nucleotide overall nucleotide contents and A3s, U3s, C3s, and G3s. The significant correlation between ENC values and whole nucleotide contents (except %G) confirmed the involvement of mutational pressure. The first and second axes values of CA were also significantly correlated with whole nucleotide content. All of the above findings indicate that mutational pressure is a significant factor influencing the codon usage bias of the N gene of NiV.

Natural selection may also alter codon usage patterns during the virus's adaptation to host cells [8, 9]. Strong correlations between Gravy values with GC3s, G3s, C3s, A3s, U3s, and ENC values were observed in current studies, indicating that viral protein characteristic has also been responsible for the observed variation in NiV codon usage. High CAI values of N gene of NiV isolates in comparison to its host (human, pig, and bat) indicated the effect of natural selection on codon usage bias. Moreover, CAI values were higher than eCAI values in respective hosts also indicated the significant adaptation of the virus to their hosts be due to natural selection.

In many RNA viruses, geographical dispersion and evolutionary processes also contribute to codon usage bias [8, 9]. In this study, geographical distribution based on CA and phylogenetic analysis were used to investigate the effects of geographical dispersion and evolutionary processes on codon usage, respectively. During CA, two different clades were formed. All Malaysian and Cambodian isolates fell into Cluster-A and Indian and Bangladeshi isolates fell into Cluster-B, while isolates from Thailand were distributed in both cluster-A and Cluster-B. This distribution of area-specific NiV isolates in specific clusters of CA graph

indicated the role of geographical distribution on codon usage bias. In the phylogenetic tree, two clades were observed and distributions of NiV isolates were similar to distributions of isolates in CA graph. Similar patterns of clustering during CA and clade formation in phylogenetic tree supported the role of evolutionary processes on codon usage bias in NiV.

The current study indicated low codon usage bias in NiV. Mutational pressure and natural selection were found to be two key factors impelling codon usage bias. In addition to mutational pressure and natural selection, geographical distribution and evolutionary processes were also influencing codon usage bias, to some extent.

Declarations

Funding Information

There is no role of any funding agencies in the current study.

Acknowledgments

We are thankful to the Dean, College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana for the support.

Disclosure of potential conflicts of interest

The authors declare that there are no conflicts of interest.

Research involving human participants and/or animals

This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Aditi, Shariff M (2019) Nipah virus infection: A review. Epidemiol Infect 147:e95. https://doi.org/10.1017/S0950268819000086
2. Looi LM, Chua KB (2007) Lessons from the Nipah virus outbreak in Malaysia. Malays. J. Pathol. 29:63–67
3. Khandia R, Singhal S, Kumar U, et al (2019) Analysis of nipah virus codon usage and adaptation to hosts. Front Microbiol 10:. https://doi.org/10.3389/fmicb.2019.00886
4. Singh RK, Dhama K, Chakraborty S, et al (2019) Nipah virus: epidemiology, pathology, immunobiology and advances in diagnosis, vaccine designing and control strategies–a comprehensive review. Vet. Q. 39:26–55
5. Martinez-Gil L, Vera-Velasco NM, Mingarro I (2017) Exploring the Human-Nipah Virus Protein-Protein Interactome. J Virol 91:. https://doi.org/10.1128/jvi.01461-17

6. Eshaghi M, Tan WS, Yusoff K (2005) Identification of epitopes in the nucleocapsid protein of Nipah virus using a linear phage-displayed random peptide library. J Med Virol 75:147–152. https://doi.org/10.1002/jmv.20249
7. Ranadheera C, Proulx R, Chaiyakul M, et al (2018) The interaction between the Nipah virus nucleocapsid protein and phosphoprotein regulates virus replication. Sci Rep 8:15994. https://doi.org/10.1038/s41598-018-34484-7
8. Singh NK, Tyagi A (2017) A detailed analysis of codon usage patterns and influencing factors in Zika virus. Arch Virol 162:1963–1973. https://doi.org/10.1007/s00705-017-3324-2
9. Singh NK, Tyagi A, Kaur R, et al (2016) Characterization of codon usage pattern and influencing factors in Japanese encephalitis virus. Virus Res 221:58–65. https://doi.org/10.1016/j.virusres.2016.05.008
10. Zhao K-N, Liu WJ, Frazer IH (2003) Codon usage bias and A+T content variation in human papillomavirus genomes. Virus Res 98:95–104. https://doi.org/10.1016/j.virusres.2003.08.019
11. Nyayanit DA, Yadav PD, Kharde R, Cherian S (2021) Natural selection plays an important role in shaping the codon usage of structural genes of the viruses belonging to the coronaviridae family. Viruses 13:. https://doi.org/10.3390/v13010003
12. Chakraborty S, Deb B, Barbhuiya PA, Uddin A (2019) Analysis of codon usage patterns and influencing factors in Nipah virus. Virus Res 263:129–138. https://doi.org/10.1016/j.virusres.2019.01.011
13. Peden JF (1999) Analysis of Codon Usage. University of Nottingham,UK.
14. Sharp PM, Li WH (1986) An evolutionary perspective on synonymous codon usage in unicellular organisms. J Mol Evol 24:28–38. https://doi.org/10.1007/BF02099948
15. Butt AM, Nasrullah I, Qamar R, Tong Y (2016) Evolution of codon usage in Zika virus genomes is host and vector specific. Emerg Microbes Infect 5:. https://doi.org/10.1038/emi.2016.106
16. Wright F (1990) The “effective number of codons” used in a gene. Gene 87:23–29. https://doi.org/10.1016/0378-1119(90)90491-9
17. Gustafsson C, Minshull J, Govindarajan S, et al (2012) Engineering genes for predictable protein expression. Protein Expr Purif 83:37–46. https://doi.org/10.1016/j.pep.2012.02.013
18. Xia X (2007) An improved implementation of codon adaptation index. Evol Bioinforma 3:53–58. https://doi.org/10.1177/117693430700300028
19. Puigbò P, Bravo IG, Garcia-Vallve S (2008) CAIcal: A combined set of tools to assess codon usage adaptation. Biol Direct 3:. https://doi.org/10.1186/1745-6150-3-38
20. Lobry JR, Gautier C (1994) Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 Escherichia coli chromosome-encoded genes. Nucleic Acids Res 22:3174–3180. https://doi.org/10.1093/nar/22.15.3174
21. Butt AM, Nasrullah I, Tong Y (2014) Genome-Wide Analysis of Codon Usage and Influencing Factors in Chikungunya Viruses. PLoS One 9:e90905. https://doi.org/10.1371/journal.pone.0090905
22. Chen Y, Shi Y, Deng H, et al (2014) Characterization of the porcine epidemic diarrhea virus codon usage bias. Infect Genet Evol 28:95–100. https://doi.org/10.1016/j.meegid.2014.09.004

23. Ahn I, Son HS (2012) Evolutionary analysis of human-origin influenza A virus (H3N2) genes associated with the codon usage patterns since 1993. Virus Genes 44:198–206. https://doi.org/10.1007/s11262-011-0687-4
24. Xu Y, Jia R, Zhang Z, et al (2015) Analysis of synonymous codon usage pattern in duck circovirus. Gene 557:138–145. https://doi.org/10.1016/j.gene.2014.12.019

Figures

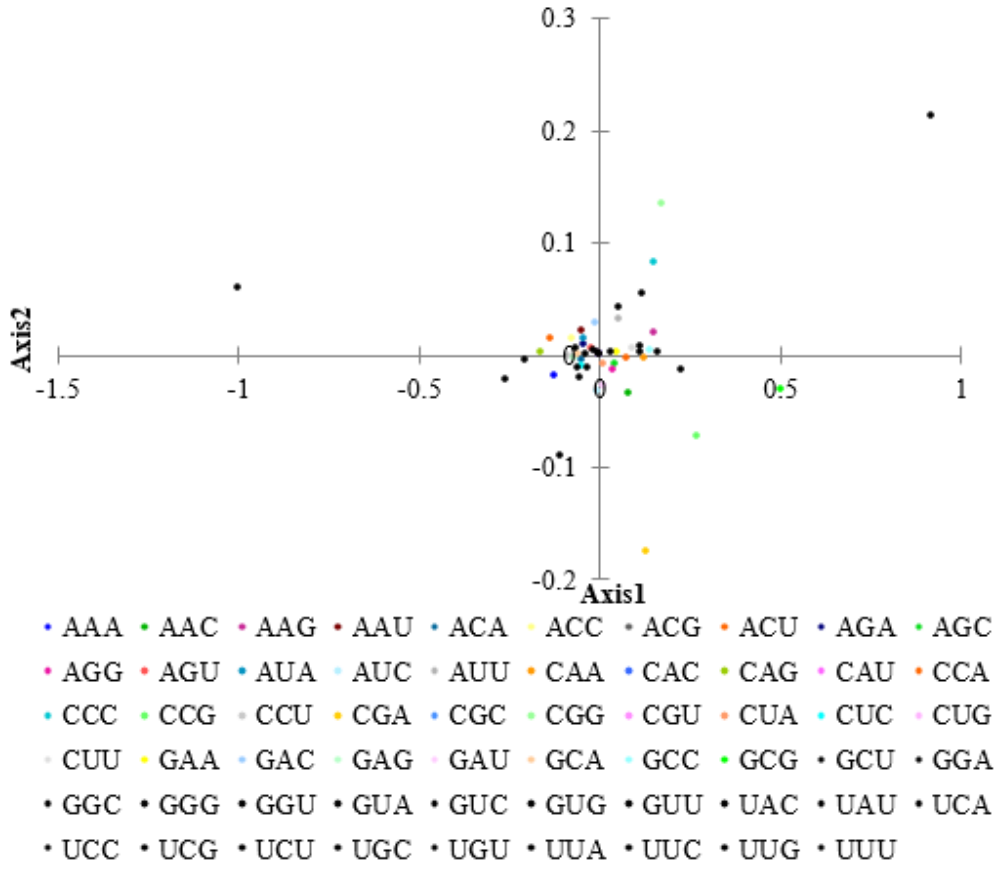

Figure 1

The correspondence analysis (CA) based on the RSCU values of N gene of NiV.

Axis 1 and 2 values for each codon, generated by CA, were represented on a scatter plot. The first axis (Axis 1) explained 86.05% of the total variance, while the second axis (Axis 2) covered 4.61% of the total variance.

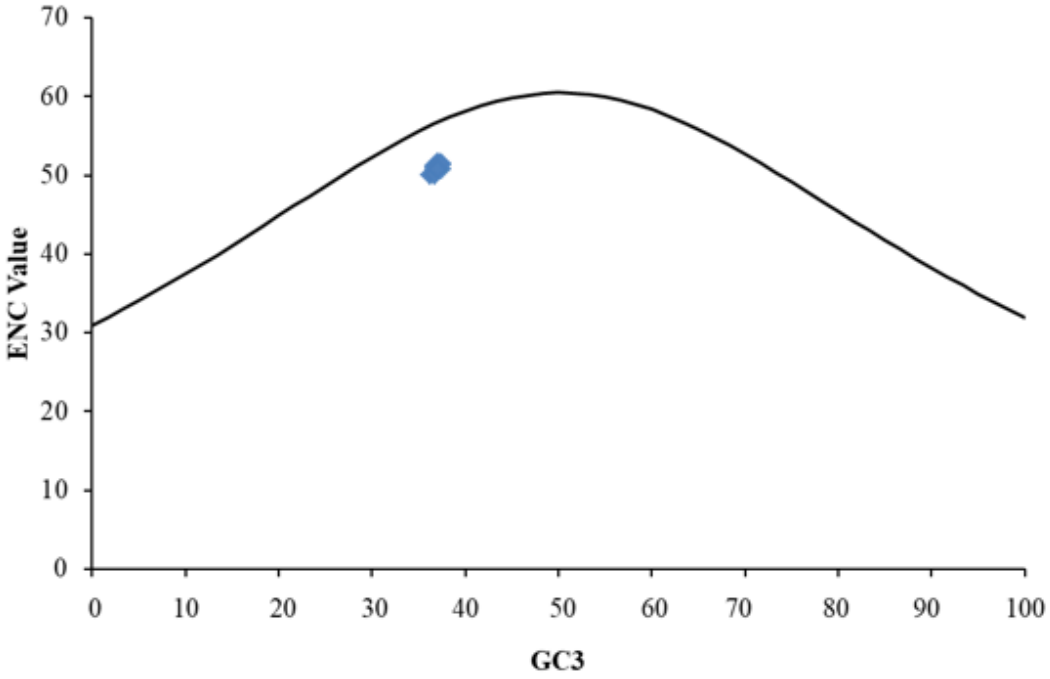

Figure 2

The relationship between the ENC values and the GC3s.

In the absence of selection, the relationship between GC3s and ENCs was shown by a solid curve line. ENC versus of N gene of various NiV isolates were clustered slightly below the expected curve indicating the presence of selection pressure on the codon usage pattern of NiV.

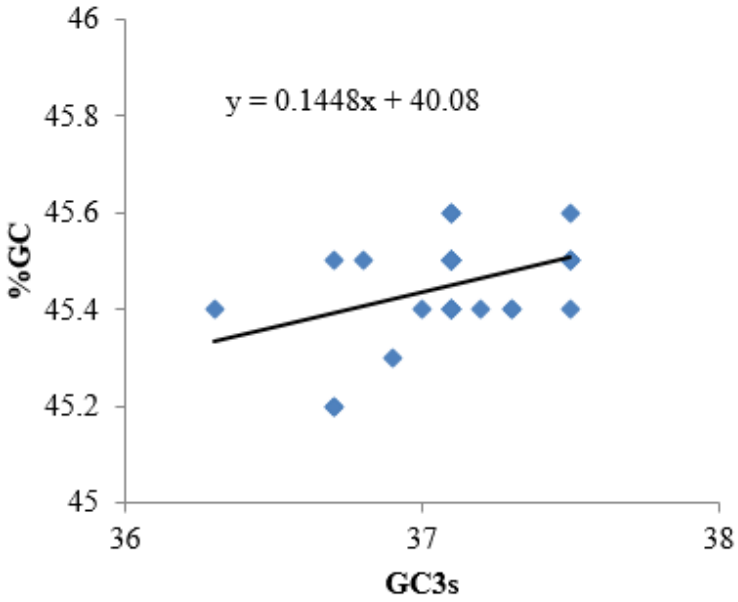

Figure 3

The Correlation analysis between % GC and GC3s.

Spearman correlation analysis between % GC and GC3s was performed ($p < 0.05$). The black solid line represents the correlation line. The correlation curve equation has also been shown on the plot.

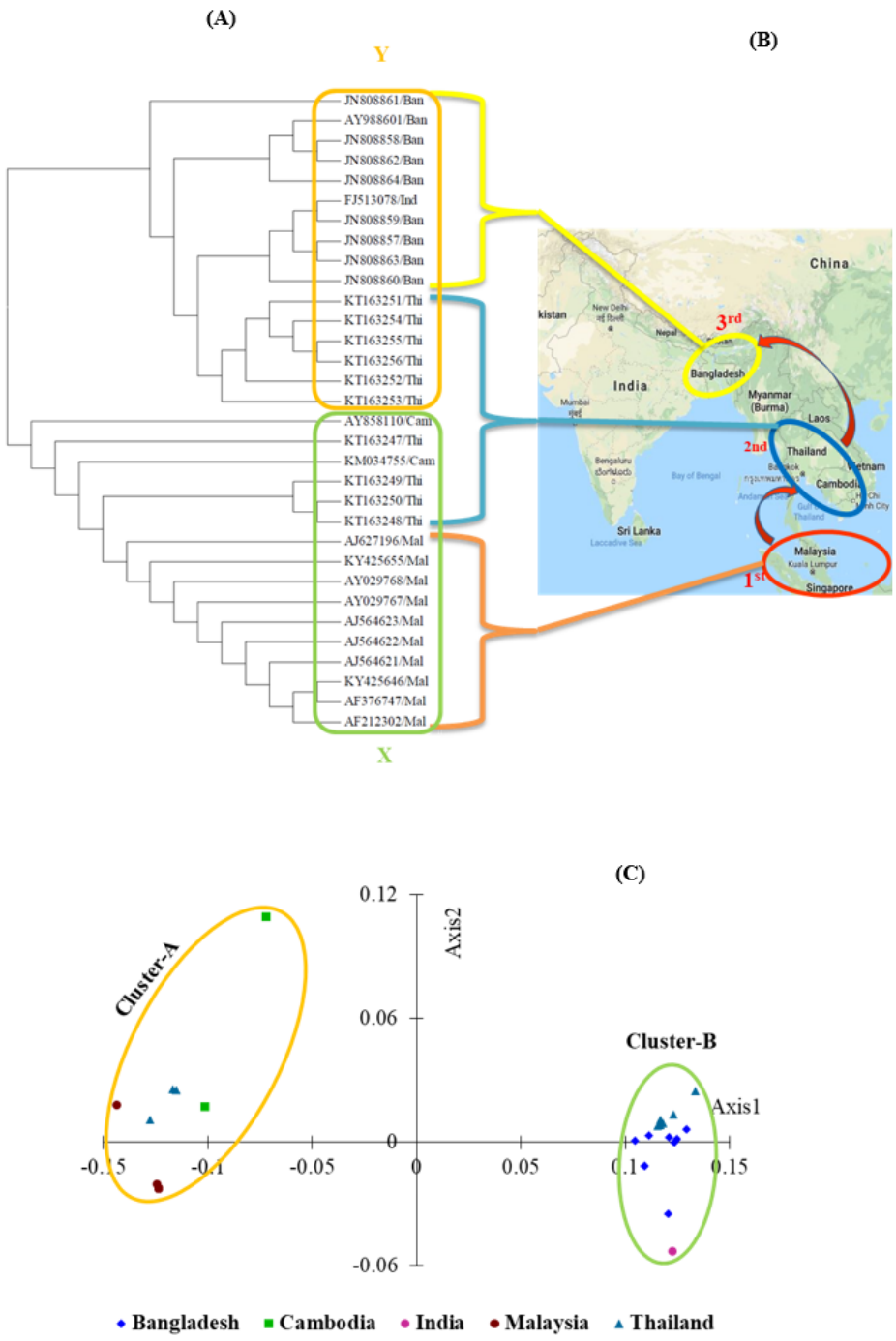

Figure 4

The phylogenetic tree based on N gene and geographical distributions of various NiV isolates.

(A) Phylogenetic tree parameters included: pairwise deletion, 1000 replicates for bootstrap analysis, neighor-joining method for tree construction. All NiV isolates were organized into two separate clades (Clade-X, and Clade-Y) (Ban, Cam, Ind, Mal, and Thi, stands for Bangladesh, Cambodia, India, Malaysia, and Thailand, respectively). (B) Based on the geographical distribution and time of N gene sequence of various isolates

reported to NCBI, NiVs are grouped into three different bunches. (C) The CA based on codon usage pattern in N gene indicated that all NiV isolates formed two clusters (Cluster-A, and Cluster-B) and viral geographical distributions played a significant role in its codon usage pattern.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementryData.docx