frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 10 July 2023
DOI 10.3389/fimmu.2023.1131647

Check for updates

OPEN ACCESS

EDITED BY
Sajjad Ahmad,
Abasyn University, Pakistan

REVIEWED BY
Samira Sanami,
Shahid Beheshti University of Medical Sciences, Iran
Hassan Alhassan,
Al Jouf University, Saudi Arabia
Dr. Bilal Shaker,
Ewha Womans University, Republic of Korea

*CORRESPONDENCE
Usman Ali Ashfaq
ashfaqua@gcuf.edu.pk

RECEIVED 25 December 2022
ACCEPTED 22 June 2023
PUBLISHED 10 July 2023

CITATION
Noor F, Ashfaq UA, Bakar A, Qasim M, Masoud MS, Alshammari A, Alharbi M and Riaz MS (2023) Identification and characterization of codon usage pattern and influencing factors in HFRS-causing hantaviruses. *Front. Immunol.* 14:1131647. doi: 10.3389/fimmu.2023.1131647

COPYRIGHT
© 2023 Noor, Ashfaq, Bakar, Qasim, Masoud, Alshammari, Alharbi and Riaz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification and characterization of codon usage pattern and influencing factors in HFRS-causing hantaviruses

Fatima Noor[1], Usman Ali Ashfaq[1]*, Abu Bakar[2], Muhammad Qasim[1], Muhammad Shareef Masoud[1], Abdulrahman Alshammari[3], Metab Alharbi[3] and Muhammad Shahid Riaz[4]

[1]Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan, [2]Centre of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, Faisalabad, Pakistan, [3]Department of Pharmacology and Toxicology, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia, [4]School of Dentistry, University of Maryland, Baltimore, MD, United States

Hemorrhagic fever with renal syndrome (HFRS) is an acute viral zoonosis carried and transmitted by infected rodents through urine, droppings, or saliva. The etiology of HFRS is complex due to the involvement of viral factors and host immune and genetic factors which hinder the development of potential therapeutic solutions for HFRS. Hantaan virus (HTNV), Dobrava-Belgrade virus (DOBV), Seoul virus (SEOV), and Puumala virus (PUUV) are predominantly found in hantaviral species that cause HFRS in patients. Despite ongoing prevention and control efforts, HFRS remains a serious economic burden worldwide. Furthermore, recent studies reported that the hantavirus nucleocapsid protein is a multi-functional protein and plays a major role in the replication cycle of the hantavirus. However, the precise mechanism of the nucleoproteins in viral pathogenesis is not completely understood. In the framework of the current study, various *in silico* approaches were employed to identify the factors influencing the codon usage pattern of hantaviral nucleoproteins. Based on the relative synonymous codon usage (RSCU) values, a comparative analysis was performed between HFRS-causing hantavirus and their hosts, suggesting that HTNV, DOBV, SEOV, and PUUV, were inclined to evolve their codon usage patterns that were comparable to those of their hosts. The results indicated that most of the overrepresented codons had AU-endings, which revealed that mutational pressure is the major force shaping codon usage patterns. However, the influence of natural selection and geographical factors cannot be ignored on viral codon usage bias. Further analysis also demonstrated that HFRS causing hantaviruses adapted host-specific codon usage patterns to sustain successful replication and transmission chains within hosts. To our knowledge, no study to date reported the factors influencing the codon usage pattern within hantaviral nucleoproteins. Thus, the proposed computational scheme can help in understanding the underlying mechanism of codon usage patterns in HFRS-causing hantaviruses which lend a helping hand in designing effective anti-HFRS treatments in future. This study, although comprehensive,

relies on in silico methods and thus necessitates experimental validation for more solid outcomes. Beyond the identified factors influencing viral behavior, there could be other yet undiscovered influences. These potential factors should be targets for further research to improve HFRS therapeutic strategies.

KEYWORDS

hemorrhagic fever with renal syndrome, hantavirus, RSCU, ENC, mutation, selection

1 Introduction

Hemorrhagic fever with renal syndrome (HFRS) has been a major notable epidemic mainly in Europe and Asia (1, 2). Annual cases of HFRS range from 60000 to 150000 and are highly correlated with rodent populations (3). The clinical picture is substantially characterized by a triad of symptoms including hypotension, high fever, kidney damage, and bleeding (4, 5). In different regions of the world, especially in Asia and Europe, HFRS is considered an acute zoonotic viral disease (2, 6, 7). The susceptible population of HFRS is heterogeneous in terms of geographical differences among several regions, space, as well as socioeconomic status (8). Since no therapeutic solutions are available to fight against the disease, therefore, supportive as well as symptomatic care that primarily focuses on regulating electrolyte balance, the pressure of blood and fluid might prevent complications in affected individuals (9, 10).

Several members of the hantavirus genus including Hantaan virus (HTNV) (11), Dobrava-Belgrade virus (DOBV) (12), Seoul virus (SEOV) (13), and Puumala virus (PUUV) (14) are the leading cause of HFRS. Hantaviruses are single-stranded, membrane-enveloped RNA viruses belonging to the family Hantaviridae within the order Bunyavirales (15–17). More importantly, research on therapeutic options requires a thorough understanding of their genome structure and the essential roles of their hantaviral proteins. The hantavirus genome is composed of three segments coding for G1 and G2 glycoproteins, nucleocapsid protein, and viral polymerase. Compared to non-pathogenic strains, pathogenic hantaviruses significantly alter the transcriptional activity of many cellular genes (18, 19). Recent studies provide solid evidence that hantaviruses' nucleocapsid proteins have a key role in virus transcription, replication, and assembly (20, 21). The nucleoprotein, encoded by the S segment, of hantaviruses consist of 429 to 433 amino acids (22). This nucleoprotein interacts with the host proteins and limits the activation of the major antiviral signaling pathways in affected cells (23).

Degeneracy or redundancy of codons offers an opportunity for evolution to increase translation productivity while retaining the same sequence of amino acids (24). The codons have undergone extensive evolution, therefore, these synonymous codons are varied among species or even organisms (25). In each genome, the frequency of synonymous codons is varied and this universal event is dubbed as codon usage bias (26, 27). Molecular evolutionary studies revealed that this pattern is pervasive among eukaryotes, prokaryotes, and viruses as well as different genes of the same organism (28). Furthermore, transcriptional factors, GC content, translation, the expression level of genes, and secondary structural motifs are the main factors driving codon usage patterns (29, 30). Transcriptional factors can influence codon preferences by binding selectively to specific DNA sequences (31). The binding of these factors promotes/suppress the usage of certain codons, impacting the translation efficiency and protein production. On the other hand, genomes with higher GC content often exhibit biased codon preferences, with certain codons enriched or depleted. This bias arises from mutational processes, as well as the selective pressure for efficient translation and protein folding (32). Despite that, the main causes of the heterogeneity in codon usage among organisms are assumed to be natural selection and mutational pressure (33, 34). A slew of studies discovered that mutational pressure, rather than selection is the primary factor determining the codon usage bias (35). Additionally, mutational pressure cannot be considered as the main driving force in the case of different RNA or DNA viruses (36). Viral genomes differ from the genomes of prokaryotes and eukaryotes in certain aspects. For example, they depend on their hosts to replicate, synthesize, and transfer protein. Therefore, it is suggested that this interaction between the virus and the host affects the viral evolution, immunological escape from the host's immune system, and viral survival (37). Host-specific selection pressures and immune-driven selection shape codon preferences in viral genes. Co-evolution between the virus and its host further impacts genetic diversity and viral adaptation (38, 39). Thus, comprehending these patterns of codon usage within viruses can help us to gain a better understanding of the genetic diversity and evolutionary dynamics in viral proteins.

This study is attributed to the identification and characterization of codon usage patterns in nucleoproteins for understanding genetic diversity and evolutionary dynamics in HTNV, DOBV, SEOV, and PUUV. Analysis of codon usage patterns would contribute to our understanding of the evolutionary mechanisms and genetic architecture of HFRS-causing hantaviruses. Thus, the present study focuses on the comprehensive analysis of genetic diversity, synonymous codon usage patterns, and factors involved in shaping the codon usage pattern of hantaviral nucleoproteins. We identified overrepresented and underrepresented codons, which could potentially be used in genetic engineering to develop better therapeutics. By leveraging the unique codon usage patterns, we

can potentially optimize gene expression, enhance protein production, and improve the efficacy of antiviral treatments. Furthermore, the comprehensive analysis of genetic diversity and synonymous codon usage patterns provides valuable insights into the evolutionary mechanisms and genetic architecture of HFRS-causing hantaviruses, paving the way for future research and the development of targeted interventions.

2 Materials and methods

2.1 Data description

Obtaining coding sequences of hantaviral nucleoproteins is a preliminary step for analyzing the codon usage bias which serves as an alternative option for unveiling the evolution of HFRS-causing hantaviruses. The coding sequence of nucleocapsid proteins of HFRS-causing hantaviruses including SEOV, HTNV, DOBV, and PUUV, were retrieved from UniProt (40) with the following UniProt identifiers: P27313 (PUUV), W0LUE3 (SEOV), Q805Q9 (DOBV), and P05133 (HTNV) respectively.

2.2 Analysis of base composition

The coding sequences of HTNV, DOBV, SEOV, and PUUV nucleoproteins were then evaluated for the identification of nucleotide components. Two independent programs named codonW (http://codonw.sourceforge.net/) (41) and CAIcal (http://genomes.urv.es/CAIcal/) (42) were employed for calculating four different types of nucleotide frequencies at the third position including A3, U3, G3, C3, frequency of G and C at first(GC1), second(GC2) and third (GC3) positions. CodonW uses several statistical tests, such as chi-squared and Fisher's exact tests, to determine if the observed codon usage frequencies deviate significantly from the expected frequencies (43, 44). Moreover, the UGA, UAG, and UAA are stop codon and does not code for amino acids. Therefore, UGA, UAG, and UAA codons are not anticipated to exhibit any codon usage bias and were thereby discarded from subsequent analysis.

2.3 Relative synonymous codon usage analysis

The Relative synonymous codon usage (RSCU) represent the codon frequency in the genome which is equal to the average frequency of all codons encoded same amino acid. CAIcal (42) software was employed for predicting RSCU within nucleoproteins of HFRS-causing viruses which in turn lends a helping hand in evaluating the codon usage bias of coding sequences. CAIcal software quantifies the similarity between the synonymous codon usage of genes and the synonymous codon frequency of a reference set and at the end measure the synonymous codon usage bias for sequence of interest (42). Codons having RSCU values above 1 represents strong bias for the corresponding codons and were referred to as 'frequent' codons. On the other hand, codons having RSCU value below than one show weak codon usage bias and were referred to as 'less-frequent' codons (45). Similarly, underrepresented and overrepresented codons were also calculated based on the RSCU value of > 1.6 and < 0.6 respectively (46). Lastly, codons having RSCU value are equal to one, representing the no codon usage bias. RSCU analysis assists in deciphering that how selection pressure and mutation shape the genetic evolution of the virus. Thus, insights into codon bias can provide valuable knowledge on viral adaptation, survival strategies, and potentially help in the design of antiviral strategies.

2.4 Factor influencing codon usage pattern

ENC (Effective Number codons) values were then plotted against GC3s value to analyze the effect of the overall composition of nucleotides on the synonymous codon usage pattern. The position of data points relative to the standard line in the ENC-GC3s plot provides insights into the influence of selection and mutation on codon usage. In the ENC-GC3s plot, if the observed data points are on the upper portion of the standard line, then it will indicate that the selection influences the pattern. While if the points are on the bottom part of the standard line, then it will indicate that mutation affects the pattern. So to uncover the interrelationship among ENC and GC3s values, the ENC values for codons at third positions (GC3s) were computed using the method introduced by Singh et al. (47) as follows:

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

Where s indicates the composition of GC3s. Further, the GC12 values (average of codons at GC1 and GC2) were then plotted against the GC3 values (48) to explore the effect of both variation and selection on the codon usage pattern of hantaviral nucleoproteins. GC12-GC3 plot examines the correlation between GC content at the first and third codon positions. A positive correlation suggests a similar pattern of variation in GC content at both positions, indicating the role of mutational bias. On the other hand, a negative correlation suggests a contrasting pattern, which can be attributed to selection.

2.5 Codon adaptation index analysis

The codon adaptation index (CAI) accurately predicts the potential expression levels of nucleoproteins relative to the codon usage pattern of expressed reference proteins (49). CAI measures the translation efficiency and nowadays it is used to manipulate the nucleotide sequences for maximal production of proteins to increase the process of vaccine designing. Regarding this, the CAIcal server (42) was employed to assess the CAI of HTNV, DOBV, SEOV, and PUUV nucleoproteins concerning the *Homo sapiens* host. The CAIcal server calculates the CAI value for a given gene based on the frequency of occurrence of synonymous codons for each amino acid in the gene sequence, and compares these

frequencies to the reference set of highly expressed genes (50). The CAI value varies from 0 and 1 which reflects the higher levels of viral adaptation with their respective host environment.

2.6 Relative codon deoptimization index analysis

Recently, Mueller et al. (51) introduced the Relative codon deoptimization index (RCDI) for comparing the codon similarities between hantaviral nucleoproteins as well as host organisms. RCDI analysis provides an ultimate estimate translation rate with their respective host organism. If the pattern in both nucleoproteins and host is found to be similar, then the RCDI value is near to one which indicates excellent adaptability with their hosts. The RCDI value of nucleoproteins was computed using CAIcal (42) for finding similarities among the codon usage pattern of HTNV, DOBV, SEOV, and PUUV with *Homo sapiens*.

2.7 Correspondence analysis

Correspondence analysis is a multivariate statistical technique for the identification of several different variations in synonymous codon usage among proteins of interest and provides a way to figure out highly expressed genes (39). Correspondence analysis was conducted on the RSCU values to minimize the effect of inconsistencies in the overall composition of amino acids. Correspondence analysis displays a set of columns and rows in a particular dataset. The RSCU values of 59 amino acids encoding nucleoproteins were computed by representing these amino acids into 59-dimensional vectors. Each dimension in 59-dimensional vectors corresponds to the RSCU values of particular amino acids. The desired result of the correspondence analysis plot indicates the discrepancies in the usage of preferred codons among coding sequences. Additionally, by producing visual outputs, it reduces unnecessary noise and complex data structures.

2.8 Correlation analysis

Spearman's rank correlation is a non-parametric measure that does not assume any particular distribution of the data and is robust to outliers. It is commonly used in codon usage analysis because it can reveal relationships between codon usage patterns and other factors that may not be linear. Spearman's rank correlation method was employed in the current study to analyze the correlation among base contents of HTNV, DOBV, SEOV, and PUUV. If the value of the correlation coefficient (r) is 1 or near 1, then the selection is the main factor affecting the pattern of codon usage in hantaviral nucleoproteins. If the correlation among base contents is significant, the mutational pressure affects the pattern. Following that, R corrplot package was employed to conduct the correlation analysis while the codonW (41) software was employed to obtain the codon usage bias and other related indicators. R corrplot package calculate correlation coefficients and perform statistical tests to assess the significance of correlations (52). Lastly, the spearman's rank correlation matrix was generated in R using the corrplot library which assists in the identification of the relationship between the nucleotide composition and the codon usage pattern.

2.9 Minimum free energy analysis

Minimum free energy was used as a measure of secondary structure formation. During the translation process, energy is released which is calculated in kcal/mol (53). A negative sign in the free energy values represents high absolute values which in turn implies a huge amount of energy loss by mRNA while adopting a stable conformation. Regarding this, RNAfold webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) (54) was used to calculate minimum free energies in hantaviral nucleoproteins. RNAfold webserver uses dynamic programming algorithm to calculate the minimum free energy of RNA secondary structures. The algorithm is based on the principles of thermodynamics and statistical mechanics, and takes into account the various energetic contributions of different base-pairing interactions in the RNA molecule (55). Later, the correlation between minimum free energy and ENC along with GC contents was then calculated to analyze the relationship among variables. In our study, we examined the correlation between minimum free energy and other variables such as ENC and GC contents. This analysis helps us explore the relationship between the stability of the secondary structure and codon usage patterns, as well as the influence of GC content on the folding potential of hantaviral nucleoproteins.

3 Results

3.1 Nucleotide compositional analysis

The coding sequence of HTNV, DOBV, SEOV, and PUUV nucleoproteins was retrieved from NCBI. The composition of nucleotides in HTNV, DOBV, SEOV, and PUUV was predicted for unveiling the possible effect of nucleotide constraints on the pattern of codon usage (Figure 1). The percentage of A, C, T, and G was 32.2% ± 0.96, 19.5% ± 1.16, 23.7% ± 1.67, and 24.6% ± 1.7 (average ± Standard Deviation (SD)) respectively, in HFRS-causing hantaviruses. These results indicate that nucleotide A of HTNV, DOBV, SEOV, and PUUV is utilized more frequently. To get a deeper insight into the nucleotide composition in HTNV, DOBV, SEOV, and PUUV coding sequences, the mean values of the codon at the third position were further analyzed. The percentage of A3, C3, T3, and G3 was 30.7% ± 1.63, 16.8% ± 2.21, 26.6% ± 2.34, and 25.7% ± 2.74 respectively. The mean and SD of AU% and GC% contents were then calculated to be 55.8% ± 2.61 and 44.11% ± 2.60 respectively, which indicates that coding sequences of HTNV, DOBV, SEOV, and PUUV are enriched with AU% as compared to GC% nucleotides. Similarly, the mean and SD of AU3% and GC3% contents were 57.39% ± 3.98 and 42.60% ± 3.97 respectively. The composition of GC content is another important indicator to

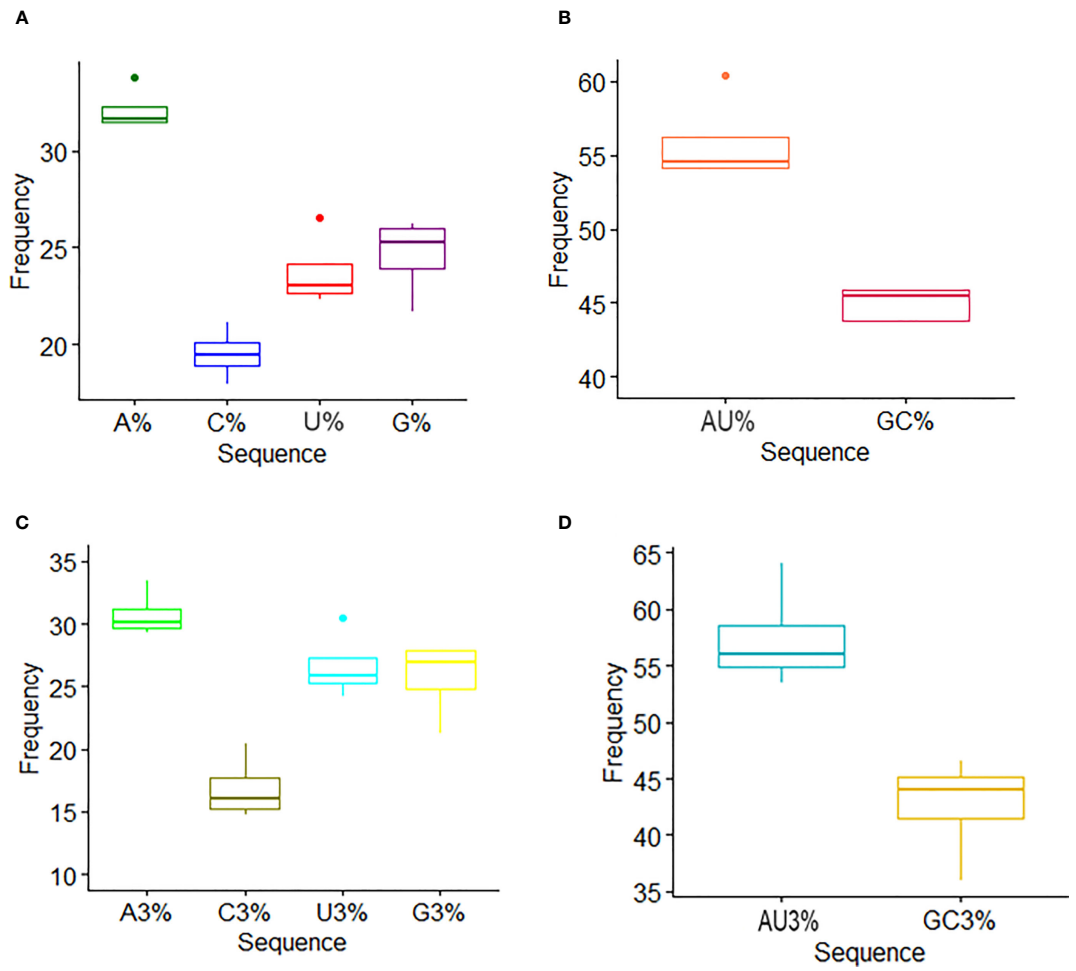

FIGURE 1
Nucleotide composition analysis. **(A)** The mean frequency for A, U, G, and C composition in nucleoprotein sequences is shown. **(B)** The mean frequency for AU and GC composition indicates AU richness. **(C)** The mean values of the nucleotide content frequency at the third codon position. **(D)** Analysis for AU and GC composition at the third codon position suggests higher AU content than GC at the third codon position.

calculate the nucleotide composition bias. The average and SD of GC content at the first composition were, 51.6% ± 2.31, at the second position the average and SD were, 38.12 ± 1.80, respectively. These findings indicate that AU-ended codons are preferred at the third position. Thus, compositional constraints have a key role in the overall composition of nucleotides in the coding sequence of HTNV, DOBV, SEOV, and PUUV.

3.2 Analysis of relative synonymous codon usage

The RSCU value of HTNV, DOBV, SEOV, and PUUV was calculated which were then compared with the RSCU value of the host (Table 1). RSCU analysis revealed that 18 amino acids named UUU (Phe), UUA (Leu), AAA (Lys), GGU (Gly), CGU (Arg), UGU (Cys), GAA (Glu), GAU (Asp), AAU (Asn), CAA (Glu), UAU (Tyr), CAU (His), UCU (Ser), CCU (Pro), ACU (Thr), GCU (Ala), GUU (Val), and AUU (Ile) ended on U and A and these18 codon occupied most region of HTNV, DOBV, SEOV, and PUUV. Recent studies on equine influenza virus, avian rotaviruses, and Crimean–Congo hemorrhagic fever virus (56–58) uncovered that viruses genomes are enriched with A and U-ended codons. Additionally, the overrepresented and underrepresented codons were analyzed based on the RSCU value. Interestingly, the nucleoproteins sequences of HTNV, DOBV, SEOV, and PUUV coding sequences contained rare over-represented sequences. Only A/U-ended codons were found to be overrepresented (RSCU value > 0.6) while G/C-ended codons were found to be under-represented which unveiled that mutational pressure influences the pattern in HTNV, DOBV, SEOV, and PUUV. In the case of HTNV and SEOV, threonine has the highest RSCU value, proline has the highest RSCU value in PUUV, and alanine has the highest RSCU value in DOBV. Furthermore, the most preferred (>1.6) as well as avoided (<0.6) codons were also highlighted. Codons with an RSCU value above 1.6 were considered preferred, while those with an RSCU value below 0.6 were considered avoided. Preferred codons are those that have a higher RSCU value, indicating that they are used more frequently than other synonymous codons for the same amino acid. Conversely, avoided codons have a lower RSCU value,

10.3389/fimmu.2023.1131647

TABLE 1 RSCU value of amino acids and corresponding synonymous codons in HTNV, DOBV, SEOV, and PUUV.

Amino acids	Codon	HTNV	DOBV	SEOV	PUUV
Phenylalanine	UUU	1.45	1.08	0.92	1.04
	UUC	0.55	0.92	1.08	0.96
Leucine	UUA	1.4	0.67	0.86	1.43
	UUG	0.98	0.93	0.69	0.78
	CUG	1.12	1.07	1.54	0.78
	CUA	0.56	0.93	0.86	0.78
	CUU	1.12	1.33	1.37	1.43
	CUC	0.84	1.07	0.69	0.78
Isoleucine	AUU	1.4	1.39	0.93	1.59
	AUC	1	0.86	1.14	0.24
	AUA	0.6	0.75	0.93	1.16
Valine	GUU	1.22	1	0.87	1.53
	GUC	0.87	0.6	1.04	0.35
	GUA	0.35	0	0.7	1.18
	GUG	1.57	2.4	1.39	0.94
Serine	UCU	0.64	2.07	0.75	1.03
	UCC	1.07	0.21	0.19	1.24
	UCA	1.93	1.86	2.81	2.07
	UCG	0.21	0	0.38	0.21
	AGU	0.86	1.03	1.12	0.83
	AGC	1.29	0.83	0.75	0.62
Proline	CCU	1.68	1.4	2	1.08
	CCC	0	0.4	0.2	0.15
	CCA	2.11	1.8	1.4	2.62
	CCG	0.21	0.4	0.4	0.15
Threonine	ACU	0.67	0.95	1.05	1.3
	ACC	0.33	1.33	0	0.32
	ACA	3	1.52	2.95	2.16
	ACG	0	0.19	0	0.22
Alanine	GCU	0.71	0.73	1.22	0.8
	GCC	0.35	0.48	0.67	0.93
	GCA	2.94	2.55	2.11	2.27
	GCG	0	0.24	0	0
Tyrosine	UAU	1.2	0.5	1.4	1.62
	UAC	0.8	1.5	0.6	0.38
Histidine	CAU	1.67	1.14	1.71	1.25
	CAC	0.33	0.86	0.29	0.75
Glutamine	CAA	0.78	1.09	1.04	1.15
	CAG	1.22	0.91	0.96	0.85

(Continued)

10.3389/fimmu.2023.1131647

TABLE 1 Continued

Amino acids	Codon	HTNV	DOBV	SEOV	PUUV
Asparagine	AAU	0.92	1	0.92	1.57
	AAC	1.08	1	1.08	0.43
Lysine	AAA	0.83	0.94	0.48	0.94
	AAG	1.17	1.06	1.52	1.06
Aspartic acid	GAU	1.43	0.97	1.13	1.32
	GAC	0.57	1.03	0.87	0.68
Glutamic acids	GAA	0.8	1	0.96	1.07
	GAG	1.2	1	1.04	0.93
Cysteine	UGU	1.6	0.4	1.2	1.43
	UGC	0.4	1.6	0.8	0.57
Arginine	AGA	1.07	1.68	2.14	2.06
	CGA	0.64	0.24	0.21	1.03
	CGC	0.21	0.24	0.86	0.69
	CGU	0.43	0.48	0.43	0.17
	AGG	2.36	2.16	1.93	1.37
Glycine	GGU	1.23	0.48	0.74	0.71
	GGC	0.62	0.96	0.3	1
	GGA	0.77	1.6	0.89	1
	GGG	1.38	0.96	2.07	1.29

The value in yellow color represents high RSCU value (>1.6) and red color indicates low RSCU value (<0.6).

indicating that they are used less frequently than other synonymous codons for the same amino acid. CCC, CCG, ACG, GCG, and GCU codons were found to be avoided in HTNV, DOBV, SEOV, and PUUV. Most of the avoided codons were ended on G/C nucleotides the preferred codons were identified as A/U ended. These findings revealed that compositional pressure of nucleotides is not considered a key criterion for characterization of codon usage bias, as the RSCU value also highlighted the codon usage pattern which in turn might uncover codon usage variation in nucleoproteins of HTNV, DOBV, SEOV, and PUUV.

3.3 Factor influencing codon usage pattern

To further evaluate the magnitude of codon usage patterns in HTNV, DOBV, SEOV, and PUUV, the ENC value GC3 plot was constructed. The ENC value ranges from 20 to 60. A lower ENC value represents a strong preference for codon usage while a higher ENC indicates a weaker preference for codon usage. ENC-GC3 plot helps in determining whether the pattern of codon usage in hantaviral nucleoproteins differs from that of the comparable synonymous codons. Further, the ENC value varied from 34.42 to 42.73 which represents several different trends in the codon usage pattern of HTNV, DOBV, SEOV, and PUUV nucleoproteins. Figure 2 represents that all dots lie just below the ENC value of 60, which demonstrated that despite of selection, the variations also affect the pattern in HTNV, DOBV, SEOV, and PUUV.

To further analyze the natural selection as well as the mutational pressure, the GC12 were plotted against GC3 to construct the neutrality plot (Figure 3). In neutrality plot, each dots represent a particular virus which corresponds to the composition of GC3 and GC13 in the neutrality plot. Furthermore, if dots fall along the diagonal distribution, it reflects zero or very little pressure of external selection, and there is no discernible difference at any of the three codon positions. The variation correlation between GC12 and GC3 is instead very low if the regression curve has a tendency to be inclined or parallel to the horizontal axis. These findings of the current study revealed the correlation between GC3 and GC12 is not statistically significant (p-value = 0.1008) which does not meet the criteria of p-value > 0.05, indicating that natural selection along with mutation pressure influences the codon usage pattern of HTNV, DOBV, SEOV, and PUUV.

3.4 Comparative host adaptability of HFRS-causing hantaviruses

The host-pathogen relationship is the extreme case of animal association and is particularly affected by different interacting factors. The viruses co-evolved with their host and are pure

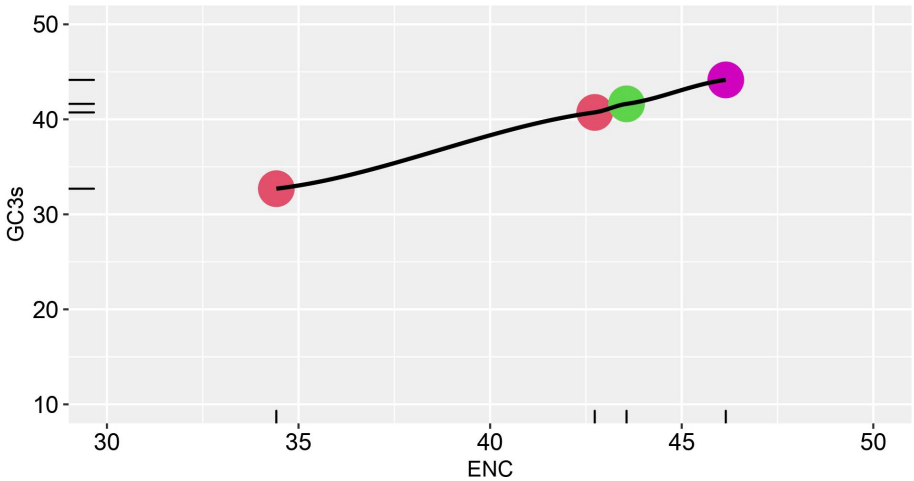

FIGURE 2
ENC-GC3 plot. The effective number of codons (ENC) plotted against the GC3S, the GC content of synonymous codons at the third position Different strains are shown in various color schemes.

parasites. A particular portion of the virus genome is associated with host-specific variations which may change the codon usage pattern of the host. Similar codon use bias is frequently seen in highly expressed viral proteins and target host proteins. So, it could be fascinating to observe how certain human parasites, like HFRS-causing hantaviruses, adapt to their hosts by reaping the benefits of the host codon usage pattern. Regarding this, the CAI values of HTNV, DOBV, SEOV, and PUUV were calculated by using *Homo sapiens* as a reference genome. The CAI value ranges from 0 to 1, with higher values indicating a better match between the codon usage of the gene and the host organism. A CAI value of 1 indicates perfect adaptation to the host organism's codon usage, while a value of 0 indicates no adaptation. Here, the CAI value demonstrated a diverse adaptability pattern among HTNV, DOBV, SEOV, and PUUV. PUUV has a low CAI score (0.719) which is closer to SEOV (0.737), followed by HTNV (0.749), and DOBV (0.75). This phenomenon suggests that, compared to other HFRS-causing hantaviruses, DOBV and HTNV have a higher level of host adaptability.

3.5 Relative codon deoptimization index analysis

The coding sequences of nucleoproteins were then compared with the reference genome to figure out the similarity between the codon usage pattern of the virus with their respective hosts. Regarding this, the RCDI values were computed for comparing the codon usage pattern of HTNV, DOBV, SEOV, and PUUV with *Homo sapiens*. A higher value of RCDI represents less adaptation to the host while low RCDI values reflect higher host adaptability. RCDI value of 1 and greater than 1 represents a high translation rate

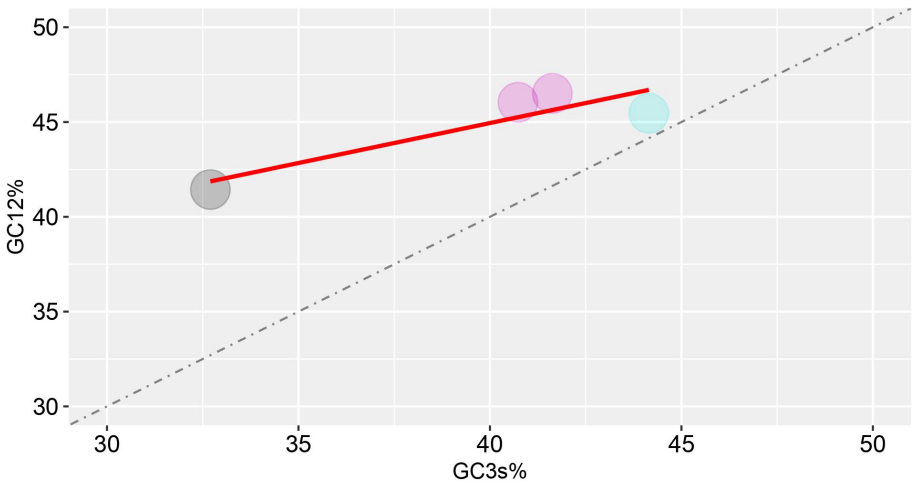

FIGURE 3
Neutrality plot analysis (GC12 vs. GC3) for the hantaviruses nucleoprotein. GC12 indicates the average value of GC contents at the first and second codon positions (GC1 and GC2), while GC3 refers to the GC contents at the third codon position.

which ultimately highlights a better adaptation to the host. The findings of the current study revealed the HTNV (1.367), DOBV (1.235), SEOV (1.334), and PUUV (1.379) had a host-adapted codon usage pattern as the RCDI values are equal to 1.

3.6 Correspondence analysis

Correspondence analysis is an effective and popular multivariate statistical analysis and it involves the plotting of codons in a 59-dimensional hyperspace in accordance with how frequently it uses each of 59 codons (removing start, Tryptophan, and termination codons). Correspondence analysis was conducted on RSCU values of HTNV, DOBV, SEOV, and PUUV to explore, interpret, and visualize the data. Correspondence analysis minimizes the dataset dimensions for huge multidimensional variables to visualize the different variables efficiently and effectively. HTNV, DOBV, SEOV, and PUUV were found to be dispersed using correspondence analysis (Figure 4). The distance among dots represents both similarities as well as dissimilarities in the pattern. The dimensions in correspondence analysis reveal the different sources of variations between a set of the multivariate data point. It is worth noting that Dimension 1 accounts for 47.9% and Dimension 2 accounts for 29.4% variations across each strain. HTNV and SEOV were found to be more closer to each other as compared to DOBV and PUUV. Altogether, these strains do not make a cluster with each other highlighting the distinction between HTNV, DOBV, SEOV, and PUUV.

3.7 Main constraints of the codon usage pattern

Mutational pressure along with selection are the two evolutionary factors which are primarily responsible for variations in the codon usage patterns. Correlation analysis of nucleotide compositions can help in the identification of primary forces causing the bias in the pattern. The overall nucleotides content (A %, U%, C%, and G%) was then correlated with the nucleotide content at the third position (A3%, U3%, C3%, and G3%) by

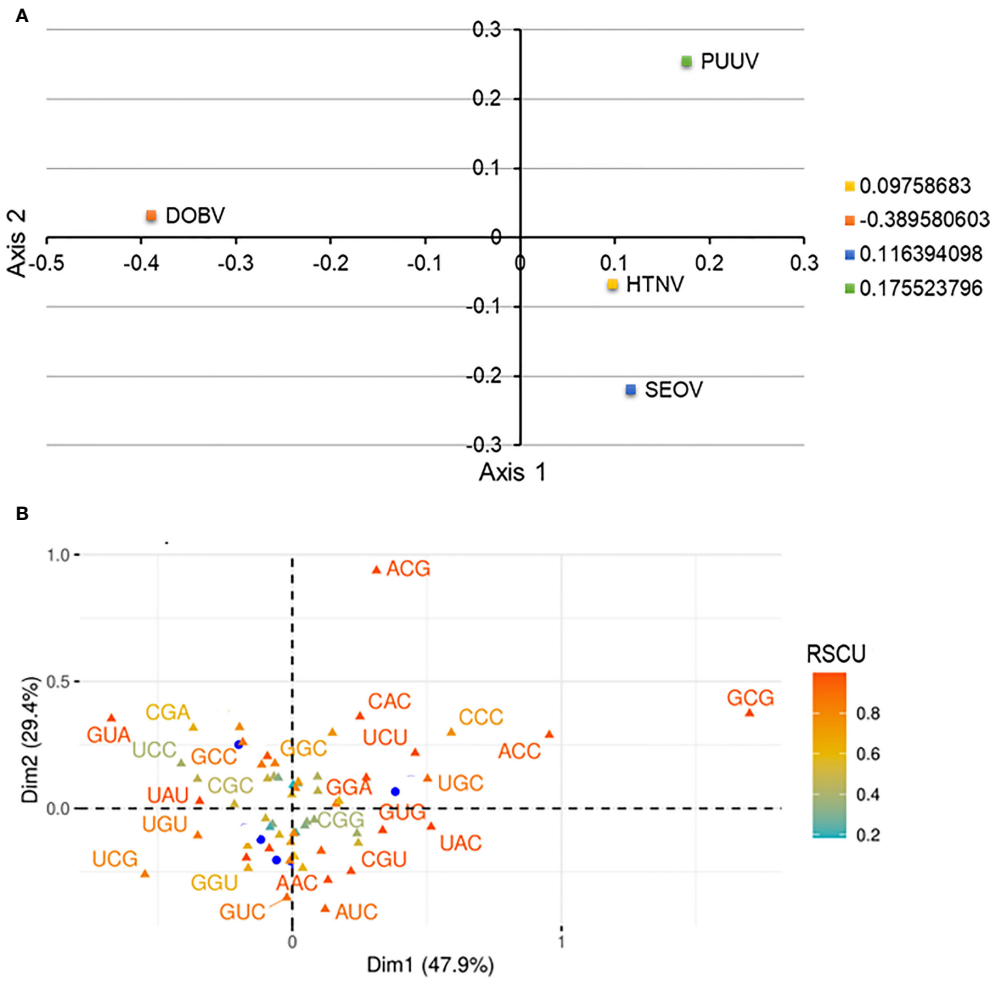

FIGURE 4
(A) Correspondence analysis of relative synonymous codon usage (RSCU) values of HTNV, DOBV, SEOV, and PUUV. The yellow square indicates HTNV, the blue square indicated SEOV, while the green and orange square indicates PUUV and DOBV, respectively. **(B)** The distribution of synonymous codons is shown along the first and second axes of the correspondence analysis.

Spearman's rank correlation (Figure 5). A significant correlation was predicted at $p < 0.01$ and $p < 0.05$, representing mutational pressure as a strong force in distinguishing the codon usage bias in the nucleoproteins of HFRS-causing hantaviruses. These findings validate that apart from natural selection and mutational pressure, nucleotide composition also plays a major role in determining the shape of pattern hantaviral nucleoproteins. Further, significant correlation (correlation coefficient= 0.8, $p < 0.01$) was predicted among ENC and CAI values of HTNV, DOBV, SEOV, and PUUV indicating that the codon usage pattern of HFRS-causing hantaviruses is limited by both mutational pressure and natural selection.

3.8 Role of minimum free energy

The minimum free energy of HTNV, DOBV, SEOV, and PUUV was - 369.20, -341.50, -369.50, and -423.70 kcal/mol, respectively. The negative sign in minimum free energy represents the loss of energy during the transcription and ultimately leads to more stable conformation. A less stable structure can be formed as a result of a higher energy release, which in turn influence the translation process. Later, the minimum free energy of nucleoproteins was then correlated using Karl Pearson's method with the ENC value of HTNV, DOBV, SEOV, and PUUV and found a significant correlation (p-value = 0.008278, r = 0.99) indicating that loss of minimum free energy was strongly related to codon usage bias. Further, the correlation between the minimum free energy of nucleoproteins and the ENC value of HTNV, DOBV, SEOV, and PUUV was calculated by Karl Pearson's method (Table 2). Correlation analysis revealed that a significant correlation exists between minimum free energy and GC%, GC1%, and GC3% of HTNV, DOBV, SEOV, and PUUV nucleoproteins. On the other hand, non-significant correlations were found among minimum free energy and GC2% and GC12% of HTNV, DOBV, SEOV, and PUUV nucleoproteins. Thus, these findings indicate that minimum free energy can affect the GC compositions and that the conformation/stability of mRNA transcripts could be related to GC constraints. It is also worth noting that, one way that minimum free energy can influence codon usage bias is through its effect on mRNA stability. Stable mRNA structures can hinder the accessibility of ribosomes to the start codon and slow down translation initiation. In contrast, less stable mRNA structures are more accessible to ribosomes and promote faster translation initiation. Therefore, genes that need to be translated quickly may have a bias towards codons that promote less stable RNA structures to enhance translation efficiency. In summary, minimum free energy can affect codon usage bias by influencing mRNA stability and the speed of translation elongation. These effects can ultimately impact protein folding and function.

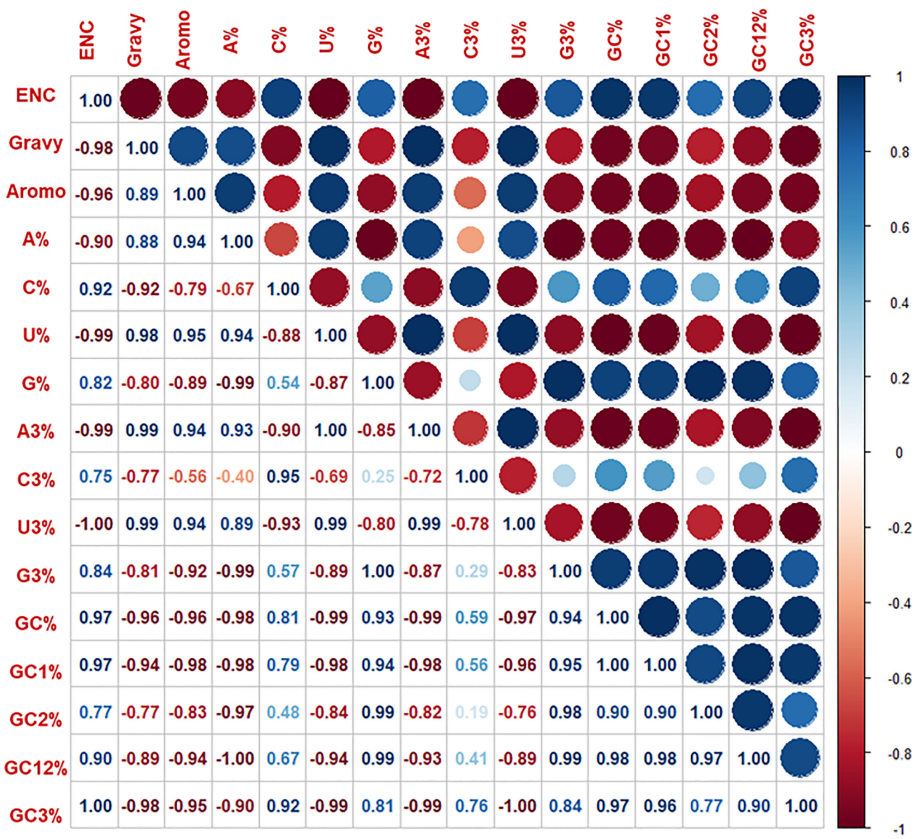

FIGURE 5
Spearman's rank correlation matrix indicated the correlation between overall base contents and base content at the third position, ENC. Gravy and aromo. The dark blue indicates a negative correlation, and the dark red indicates a positive correlation; the higher value indicates a more significant correlation.

TABLE 2 Correlation between minimum free energy and GC contents.

Correlation	GC%	GC1%	GC2%	GC3%	GC12%
Correlation-coffiecient	0.93	0.93	0.68	0.99	0.84
p-value	0.06	0.06	0.31	0.008	0.15
t-test	3.74	3.59	1.33	10.84	2.23

4 Discussion

HFRS is a group of clinically similar illnesses caused by hantaviruses belonging to family Bunyaviridae (59). The onset of the disease is characterized by various symptoms including high fever, hemorrhage, and degree of renal insufficiency, which further evolve into coagulation disorders, acute renal failure, and anemia (4, 5). A large population of mammals and rodents harboring hantaviruses lived in towns, cities, and suburbs, therefore, now efforts are being made to avoid direct contact of hantavirus-bearing organisms with humans. Currently it is estimated that 150,000 to 200,000 cases of hantavirus disease occur each year, the majority being reported in Asia (60). Despite the high morbidity and case-fatality rates of HFRS and HCPS, respectively, no vaccine or drug is currently proven to be preventive or therapeutic. Supportive therapy is the mainstay of care for patients with hantavirus infections. Care includes careful management of the patient's fluid (hydration) and electrolyte (e.g., sodium, potassium, chloride) levels, maintenance of correct oxygen and blood pressure levels, and appropriate treatment of any secondary infections (61). There is currently no effective treatment available for either HFRS. The most severe form of HFRS is caused by four different types of hantaviruses including HTNV, DOBV, SEOV, and PUUV (62). The nucleoproteins are more prevalent proteins in cells with viral infection and are highly conserved throughout the genus of hantavirus (63, 64). Thus, it is important to analyze the main forces causing variation in the function of hantaviral proteins which ultimately lend a helping hand in the development of novel treatment options for HFRS. More importantly, as the hantaviruses rely on the host cell's machinery for its replication, codon usage bias could play a role in host adaptation and the virulence of the virus.

Codon usage bias is a distinctive characteristic of many organisms and has been noted in viruses such as influenza (65, 66).Relatively, there are more reports on codon usage in genomes from viruses with great harm to mankind, such as SARS, human immunodeficiency virus, influenza virus A, and hepatitis virus (67). For instance, Khattak et al. (68) conducted a genome-wide analysis of codon usage patterns for deciphering the global heterogeneity of COVID-19. Their study proposed that the general root ancestry of the global genomes is different with different genome's level adaptations to the host. Feng et al. (69) performed a comprehensive analysis of codon usage patterns in Chinese Porcine Circoiruses (PCV) based on their major protein-coding sequences. Their findings unveiled a weak usage bias among the four PCV species and showed that in addition to mutation pressure, natural selection played a major role in PCV codon usage. Additionally, recent studies also reported that the overall pattern of codon usage in both host organisms and viruses improves the efficiency and accuracy of translation in corresponding amino acids (45, 70). While variations in the patterns ascertain the proper folding of viral nucleoproteins (71–74). Thus, analyzing codon usage patterns among HFRS-causing hantaviruses could enhance our understanding of their genetic architecture and evolutionary mechanisms. This, in turn, provide a novel insight in understanding the functioning of nucleoproteins for development of potential therapeutic strategies to fight against HFRS.

In the framework of current study, multiple systematic analytical strategies were employed to determine the factor that influences the codon usage patterns of nucleoproteins. Initially, an ENC-GC3 plot analysis was performed for analyzing the main forces behind the variation of codon usage patterns among hantaviral nucleoproteins. For this, first ENC values were calculated which were then plotted against the GC3 values. The findings proposed that ENC values are much higher as compared to ENC values of viruses causing Ebola= 57.23 (73), Crimean–Congo hemorrhagic fever= 52.34 (75), hepatitis C= 52.62) (74), and Bluetongue= 57.9) (76) indicating high bias in hantaviral nucleoproteins. A high codon usage bias affects the replication process in the host cell which increased the competition of protein synthesis in both the host and virus. Eventhough the ENC value determines the codon usage bias, these values alone does not reflect the factor that influences the codon usage pattern.

Furthermore, AT-ended codons were found to be more abundant in hantaviral nucleoproteins. Zhang et al. (77), analyzed the base content of torque tenosus virus 1 and reported A% > G % > C% > T%, suggesting the preferred use of A- over T-ending codons. Bouquet et al. (78), analyzed genetic and codon characteristics of the hepatitis E virus and reported the distribution of G and T bases to be ~ 25% each, while A was highly preferred over C. Another study conducted by Sheng et al. (67) reported the GC composition of the porcine circovirus genome to be 48.61%, playing a preferential role in synonymous codon usage. Later, to predict how these nucleotide contents altered the pattern of codon usage in nucleoproteins, we formulated our assumption from ENC–GC3 plot analysis. It indicates that mutational pressure is the main driving force contributing to the variations in the hantaviral nucleoproteins. Previous work also reported that mutational pressure influences the pattern of codon usage in Zaire ebolavirus (73). These variations are frequently constrained by the base composition at the third positions, while natural selection is mainly confined to the composition of the nucleotide at the first as well as second positions (74, 79). Thus, we conducted correlation analysis among nucleotide contents of nucleoproteins and formulated that the overall composition of

nucleotide also plays important role in shaping the codon usage patterns.

To sum up, we tried to capture the characteristics of codon usage patterns in hantaviral nucleoproteins. Consequently, our study uncovered that HFRS-causing hantaviruses had a comparatively low bias and that all over-represented codons ended with A/U. Both mutational pressures, as well as natural selection, were contributors to the bias. This is the first study in investigating the codon biology in hantaviruses. Such information does not only bring about a new perspective for understanding the mechanisms of biased usage of synonymous codons but also provide useful clues for evolutionary studies. In future, our study will pave the way for designing groundbreaking approaches such as codon pair deoptimization to modify the expression efficiency of hantaviral nucleoproteins, and the usage of least preferred codons in nucleoproteins to reduce viral pathogenesis and virulence during the development of effective and safe vaccine candidate.

Codon pair deoptimization and the use of least preferred codons are strategies used to modulate the viral protein expression and attenuate viral replication. These approaches have shown promise in vaccine development by altering the codon usage patterns of viral genes, thereby reducing viral fitness and virulence. In the case of HFRS-causing hantaviruses, applying codon pair deoptimization or least preferred codon strategies could be explored as potential vaccine design strategies. By intentionally introducing suboptimal codons or reducing the usage of preferred codons in viral antigen genes, it may be possible to attenuate the replication and pathogenicity of hantaviruses, leading to the development of safer and more effective vaccines. Deoptimized viruses can express an antigenic repertoire of both B- and T-cell epitopes as they have the complete sequence of amino acids and successfully multiply *in vitro* while being highly attenuated *in vivo*, which is crucial for the development of an effective and putative vaccine. Thus, this work offers an entirely new outlook regarding genetic diversity which may contribute to the development of novel approaches for future research on the evolutionary model, their origin, and host adaptation of HFRS-causing hantaviruses.

5 Conclusion

Our findings proposed that analyzing variation in the codon usage pattern may offer an up-to-date strategy for unveiling the genetic diversity among hantaviral nucleoproteins. It is worth noting that mutational pressure is the main force that contributed to the variation in codon usage patterns on hantaviral nucleoproteins. The overall composition of nucleotides can also act as an influencing factor in shaping the codon usage patterns of hantaviral nucleoproteins. The current study not only provides knowledge about the variations in hantaviral nucleoproteins but also lends a helping hand in identifying the factors that drive hantavirus evolution. These findings could provide a novel way of further understanding the evolutionary changes related to viral survival, host adaptation, and virulence. The significant adaptation of hantaviral nucleoproteins to humans makes these proteins a popular indicator of evolution for HTNV, DOBV, SEOV, and PUUV host speciation, and may assist in understanding the epidemic character of HFRS in humans. The information from this research may not only be helpful to get new insights into the evolution of hantaviruses but also have potential value for developing novel vaccines to fight against HFRS. Additionally, it might improve our understanding of how nucleoproteins function in HFRS and shed light on the therapeutic options. Such findings will be conducive to understanding the elements that contribute to viral evolution and adaptation to hosts.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

FN, UAA, and MQ contributed to the conception and design of the study. AB, MSM, MA, and MSR collected the data. FN performed the analysis. FN and UAA wrote the first draft of the manuscript. AB, MQ, MSM, AA, MA, and MSR revised the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

Authors are thankful to the Researchers Supporting Project number (RSP2023R491), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. She K, Li C, Qi C, Liu T, Jia Y, Zhu Y, et al. Epidemiological characteristics and regional risk prediction of hemorrhagic fever with renal syndrome in Shandong Province, China. *Int J Environ Res Public Health* (2021) 18(16):8495. doi: 10.3390/ijerph18168495

2. Tariq M, Kim D-M. Hemorrhagic fever with renal syndrome: literature review, epidemiology, clinical picture and pathogenesis. *Int J Infect Dis* (2022) 54(1):1. doi: 10.3947/ic.2021.0148

3. Teng J, Ding S, Zhang H, Wang K, Hu X. Bayesian Spatiotemporal modelling analysis of hemorrhagic fever with renal syndrome outbreaks in China using r-inla. *Zoonoses Public Health* (2022) 70(1):46-57. doi: 10.1111/zph.12999

4. Jiang H, Du H, Wang LM, Wang PZ, Bai XF. Hemorrhagic fever with renal syndrome: pathogenesis and clinical picture. *Front Cell Infect Microbiol* (2016) 6:1. doi: 10.3389/fcimb.2016.00001

5. Noor F, Ashfaq U, Asif M, Adeel M, Alshammari A, Alharbi M. Comprehensive computational analysis reveals yxxφ [I/L/M/F/V] motif and yxxφ-like tetrapeptides across hfrs causing hantaviruses and their association with viral pathogenesis and host immune regulation. *Front Immunol* (2022) 13. doi: 10.3389/fimmu.2022.1031608

6. Moolla N, Adler D, Blumberg L, Glass A, Grobbelaar A, Le Roux C, et al. First report of an imported case of haemorrhagic fever with renal syndrome in South Africa. *S Afr Med J* (2022) 112(8):519–21. doi: 10.7196/SAMJ.2022.v112i8.16540

7. Oliveira SVD, Faccini-Martínez ÁA. Hantavirus infection and the renal syndrome. *Trop Nephrology. Springer* (2020) . p:175–92. doi: 10.1007/978-3-030-44500-3_14

8. Xiao H, Tong X, Gao L, Hu S, Tan H, Huang ZY, et al. Spatial heterogeneity of hemorrhagic fever with renal syndrome is driven by environmental factors and rodent community composition. *PloS Negl Trop Dis* (2018) 12(10):e0006881. doi: 10.1371/journal.pntd.0006881

9. Cho S, Kim W-K, No JS, Lee S-H, Jung J, Yi Y, et al. Urinary genome detection and tracking of hantaan virus from hemorrhagic fever with renal syndrome patients using multiplex pcr-based next-generation sequencing. *PloS Negl Trop Dis* (2021) 15 (9):e0009707. doi: 10.1371/journal.pntd.0009707

10. Yang X, Wang C, Wu L, Jiang X, Zhang S, Jing F. Hemorrhagic fever with renal syndrome with secondary hemophagocytic lymphohistiocytosis in West China: a case report. *BMC Infect Dis* (2019) 19(1):1–4. doi: 10.1186/s12879-019-4122-0

11. Lee HW, Baek LJ, Johnson KM. Isolation of hantaan virus, the etiologic agent of Korean hemorrhagic fever, from wild urban rats. *J Infect Dis* (1982) 146(5):638–44. doi: 10.1093/infdis/146.5.638

12. Papa A. Dobrava-Belgrade virus: phylogeny, epidemiology, disease. *Antiviral Res* (2012) 95(2):104–17. doi: 10.1016/j.antiviral.2012.05.011

13. Lee HW, Lee PW, Johnson KM. Isolation of the etiologic agent of Korean hemorrhagic fever. *J Infect Dis* (1978) 137(3):298–308. doi: 10.1093/infdis/137.3.298

14. Sironen T, Vaheri A, Plyusnin A. Molecular evolution of puumala hantavirus. *J Virol* (2001) 75(23):11803–10. doi: 10.1128/JVI.75.23.11803-11810.2001

15. Liu R, Ma H, Shu J, Zhang Q, Han M, Liu Z, et al. Vaccines and therapeutics against hantaviruses. *Front Microbiol* (2020) 10:2989. doi: 10.3389/fmicb.2019.02989

16. Milholland MT, Castro-Arellano I, Suzán G, Garcia-Peña GE, Lee TE, Rohde RE, et al. Global diversity and distribution of hantaviruses and their hosts. *EcoHealth* (2018) 15(1):163–208. doi: 10.1007/s10393-017-1305-2

17. Klempa B. Reassortment events in the evolution of hantaviruses. *Virus Genes* (2018) 54(5):638–46. doi: 10.1007/s11262-018-1590-z

18. Kaukinen P, Vaheri A, Plyusnin A. Hantavirus nucleocapsid protein: a multifunctional molecule with both housekeeping and ambassadorial duties. *Arch Virol* (2005) 150(9):1693–713. doi: 10.1007/s00705-005-0555-4

19. Reuter M, Krüger DH. The nucleocapsid protein of hantaviruses: much more than a genome-wrapping protein. *Virus Genes* (2018) 54(1):5–16. doi: 10.1007/s11262-017-1522-3

20. Muyangwa M, Martynova EV, Khaiboullina SF, Morzunov SP, Rizvanov AA. Hantaviral proteins: structure, functions, and role in hantavirus infection. *Front Microbiol* (2015) 6:1326. doi: 10.3389/fmicb.2015.01326

21. Hussein IT, Mir MA. How hantaviruses modulate cellular pathways for efficient replication? *Front Biosci (Elite Ed)* (2013) 5(1):154–66. doi: 10.2741/E604

22. Alminaite A, Halttunen V, Kumar V, Vaheri A, Holm L, Plyusnin A. Oligomerization of hantavirus nucleocapsid protein: analysis of the n-terminal coiled-coil domain. *J Virol* (2006) 80(18):9073–81. doi: 10.1128/JVI.00515-06

23. Noor F, Ashfaq UA, Bakar A, Ul Haq W, Allemailem KS, Alharbi BF, et al. Discovering common pathogenic processes between covid-19 and hfrs by integrating rna-seq differential expression analysis with machine learning. *Front Microbiol* (2023) 14:1188. doi: 10.3389/fmicb.2023.1175844

24. McFeely CA, Dods KK, Patel SS, Hartman MC. Expansion of the genetic code through reassignment of redundant sense codons using fully modified trna. *Nucleic Acids Res* (2022) 50(19):11374–86. doi: 10.1093/nar/gkac846

25. Sharp PM, Emery LR, Zeng K. Forces that influence the evolution of codon bias. *Philos Trans R Soc B* (2010) 365(1544):1203–12. doi: 10.1098/rstb.2009.0305

26. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. *Mol Biol Rep* (2021) 49(1):1–27. doi: 10.1007/s11033-021-06749-4

27. Behura SK, Severson DW. Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biol Rev* (2013) 88(1):49–61. doi: 10.1111/j.1469-185X.2012.00242.x

28. Zhao Y, Zheng H, Xu A, Yan D, Jiang Z, Qi Q, et al. Analysis of codon usage bias of envelope glycoprotein genes in nuclear polyhedrosis virus (Npv) and its relation to evolution. *BMC Genom* (2016) 17(1):1–10. doi: 10.1186/s12864-016-3021-7

29. Zhou Z, Dang Y, Zhou M, Li L, C-h Yu, Fu J, et al. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc Natl Acad Sci* (2016) 113(41):E6117–E25. doi: 10.1073/pnas.1606724113

30. Quax TE, Claassens NJ, Söll D, van der Oost J. Codon bias as a means to fine-tune gene expression. *Mol Cell* (2015) 59(2):149–61. doi: 10.1016/j.molcel.2015.05.035

31. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. *Mol Biol Rep* (2022) 49(1):539–65. doi: 10.1007/s11033-021-06749-4

32. Palidwor GA, Perkins TJ, Xia X. A general model of codon bias due to gc mutational bias. *PloS One* (2010) 5(10):e13431. doi: 10.1371/journal.pone.0013431

33. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* (2011) 12(1):32–42. doi: 10.1038/nrg2899

34. Sueoka N, Kawanishi Y. DNA G+ c content of the third codon position and codon usage biases of human genes. *Gene* (2000) 261(1):53–62. doi: 10.1016/S0378-1119(00)00480-7

35. Tao P, Dai L, Luo M, Tang F, Tien P, Pan Z. Analysis of synonymous codon usage in classical swine fever virus. *Virus Genes* (2009) 38(1):104–12. doi: 10.1007/s11262-008-0296-z

36. Belalov IS, Lukashev AN. Causes and implications of codon usage bias in rna viruses. *PloS One* (2013) 8(2):e56642. doi: 10.1371/journal.pone.0056642

37. Kazlauskas D, Krupovic M, Venclovas Č. The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. *Nucleic Acids Res* (2016) 44(10):4551–64. doi: 10.1093/nar/gkw322

38. Lauring AS. Within-host viral diversity: a window into viral evolution. *Annu Rev Virol* (2020) 7:63–81. doi: 10.1146/annurev-virology-010320-061642

39. Wong EH, Smith DK, Rabadan R, Peiris M, Poon LL. Codon usage bias and the evolution of influenza a viruses. codon usage biases of influenza virus. *BMC evolutionary Biol* (2010) 10(1):1–14. doi: 10.1186/1471-2148-10-253

40. Kalita P, Lyngdoh DL, Padhi AK, Shukla H, Tripathi T. Development of multi-epitope driven subunit vaccine against fasciola gigantica using immunoinformatics approach. *Int J Biol Macromol* (2019) 138:224–33. doi: 10.1016/j.ijbiomac.2019.07.024

41. Peden J, Sharp P. Codonw: correspondence analysis of codon usage. (2005). Available at: http://codonw.sourceforge.net/.

42. Puigbò P, Bravo IG, Garcia-Vallve S. Caical: a combined set of tools to assess codon usage adaptation. *Biol Direct* (2008) 3(1):1–8. doi: 10.1186/1745-6150-3-38

43. Peden JF. Analysis of codon usage (1999). PhD thesis, University of Nottingham, UK *Anal Codon Usage* (1999).

44. Jia X, Liu S, Zheng H, Li B, Qi Q, Wei L, et al. Non-uniqueness of factors constraint on the codon usage in bombyx mori. *BMC Genomics* (2015) 16:1–12. doi: 10.1186/s12864-015-1596-z

45. Butt AM, Nasrullah I, Tong Y. Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. *PloS One* (2014) 9(3):e90905. doi: 10.1371/journal.pone.0090905

46. Khandia R, Alqahtani T, Alqahtani AM. Genes common in primary immunodeficiencies and cancer display overrepresentation of codon ctg and dominant role of selection pressure in shaping codon usage. *Biomedicines* (2021) 9 (8):1001. doi: 10.3390/biomedicines9081001

47. Singh NK, Tyagi A, Kaur R, Verma R, Gupta PK. Characterization of codon usage pattern and influencing factors in Japanese encephalitis virus. *Virus Res* (2016) 221:58–65. doi: 10.1016/j.virusres.2016.05.008

48. Nasrullah I, Butt AM, Tahir S, Idrees M, Tong Y. Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on marburg virus evolution. *BMC Evol Biol* (2015) 15(1):1–15. doi: 10.1186/s12862-015-0456-4

49. Khandia R, Singhal S, Kumar U, Ansari A, Tiwari R, Dhama K, et al. Analysis of nipah virus codon usage and adaptation to hosts. *Front Microbiol* (2019) 10:886. doi: 10.3389/fmicb.2019.00886

50. Bylaiah S, Shedole S, Suresh KP, Gowda L, Patil SS, Indrabalan UB, et al. Relative analysis of codon usage and nucleotide bias between anthrax toxin genes subsist Inpxo1 plasmid of bacillus anthracis. *Ann Romanian Soc Cell Biol* (2021) 25(4):5758–74.

51. Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E. Reduction of the rate of poliovirus protein synthesis through Large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J Virol* (2006) 80 (19):9687–96. doi: 10.1128/JVI.00738-06

52. Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J. Package 'Corrplot'. *Statistician* (2017) 56(316):e24.

53. Mathews DH, Turner DH. Prediction of rna secondary structure by free energy minimization. *Curr Opin Struct Biol* (2006) 16(3):270–8. doi: 10.1016/j.sbi.2006.05.010

54. Hofacker IL. Vienna Rna secondary structure server. *Nucleic Acids Res* (2003) 31 (13):3429–31. doi: 10.1093/nar/gkg599

55. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna rna websuite. *Nucleic Acids Res* (2008) 36(suppl_2):W70–W4. doi: 10.1093/nar/gkn188

56. Kattoor JJ, Malik YS, Sasidharan A, Rajan VM, Dhama K, Ghosh S, et al. Analysis of codon usage pattern evolution in avian rotaviruses and their preferred host. *Infect Genet Evol* (2015) 34:17–25. doi: 10.1016/j.meegid.2015.06.018

57. Kumar N, Bera BC, Greenbaum BD, Bhatia S, Sood R, Selvaraj P, et al. Revelation of influencing factors in overall codon usage bias of equine influenza viruses. *PloS One* (2016) 11(4):e0154376. doi: 10.1371/journal.pone.0154376

58. Rahman SU, Yao X, Li X, Chen D, Tao SJI. Genetics, evolution. analysis of codon usage bias of Crimean-Congo hemorrhagic fever virus and its adaptation to hosts. *Infect Genet Evol* (2018) 58:1–16.

59. Shid SJ, Sapate RB, Raut ID, Dhadde GS, Mali HS, Joshi S, et al. Hantavirus infection: an overview. *Asian J Pharm Res* (2022) 12(2):150–4.

60. Khan A, Khan M, Ullah S, Wei D-Q. Hantavirus: the next pandemic we are waiting for? *Interdiscip Sci Comput Life Sci* (2021) 13:147–52. doi: 10.1007/s12539-020-00413-4

61. Peters M CJ, Simpson M, PhD MPH, Gary L, Levy M, PhD H. Spectrum of hantavirus infection: hemorrhagic fever with renal syndrome and hantavirus pulmonary syndrome. *Annu Rev Med* (1999) 50(1):531–45. doi: 10.1146/annurev.med.50.1.531

62. Schmaljohn C. Vaccines for hantaviruses. *Vaccine* (2009) 27:D61–D4. doi: 10.1016/j.vaccine.2009.07.096

63. Yoshimatsu K, Arikawa J. Serological diagnosis with recombinant n antigen for hantavirus infection. *Virus Res* (2014) 187:77–83. doi: 10.1016/j.virusres.2013.12.040

64. Amada T, Yoshimatsu K, Yasuda SP, Shimizu K, Koma T, Hayashimoto N, et al. Rapid, whole blood diagnostic test for detecting anti-hantavirus antibody in rats. *J Virol Methods* (2013) 193(1):42–9. doi: 10.1016/j.jviromet.2013.04.021

65. Kryazhimskiy S, Bazykin GA, Dushoff J. Natural selection for nucleotide usage at synonymous and nonsynonymous sites in influenza a virus genes. *J Virol* (2008) 82 (10):4938–45. doi: 10.1128/JVI.02415-07

66. Cox NJ, Subbarao K. Global epidemiology of influenza: past and present. *Annu Rev Med* (2000) 51(1):407–21. doi: 10.1146/annurev.med.51.1.407

67. Liu X-S, Zhang Y-G, Fang Y-Z, Wang Y-L. Patterns and influencing factor of synonymous codon usage in porcine circovirus. *Virology* (2012) 9:1–9.

68. Khattak S, Rauf MA, Zaman Q, Ali Y, Fatima S, Muhammad P, et al. Genome-wide analysis of codon usage patterns of Sars-Cov-2 virus reveals global heterogeneity of COVID-19. *Biomolecules* (2021) 11(6):912. doi: 10.3390/biom11060912

69. Feng H, Segalés J, Wang F, Jin Q, Wang A, Zhang G, et al. Comprehensive analysis of codon usage patterns in Chinese porcine circoviruses based on their major protein-coding sequences. *Viruses* (2022) 14(1):81. doi: 10.3390/v14010081

70. Butt AM, Nasrullah I, Qamar R, Tong Y. Evolution of codon usage in zika virus genomes is host and vector specific. *Emerg Microbes Infect* (2016) 5(1):1–14. doi: 10.1038/emi.2016.106

71. Aragonès L, Guix S, Ribes E, Bosch A, Pintó RM. Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis a virus capsid. *PloS Pathog* (2010) 6(3):e1000797.

72. Costafreda MI, Pérez-Rodriguez FJ, D'Andrea L, Guix S, Ribes E, Bosch A, et al. Hepatitis a virus adaptation to cellular shutoff is driven by dynamic adjustments of codon usage and results in the selection of populations with altered capsids. *J Virol* (2014) 88(9):5029–41. doi: 10.1128/JVI.00087-14

73. Cristina J, Moreno P, Moratorio G, Musto H. Genome-wide analysis of codon usage bias in ebolavirus. *Virus Res* (2015) 196:87–93. doi: 10.1016/j.virusres.2014.11.005

74. Hu J-S, Wang Q-Q, Zhang J, Chen H-T, Xu Z-W, Zhu L, et al. The characteristic of codon usage pattern and its evolution of hepatitis c virus. *Infect Genet Evol* (2011) 11 (8):2098–102. doi: 10.1016/j.meegid.2011.08.025

75. Rahman SU, Yao X, Li X, Chen D, Tao S. Analysis of codon usage bias of Crimean-Congo hemorrhagic fever virus and its adaptation to hosts. *Infect Genet Evol* (2018) 58:1–16. doi: 10.1016/j.meegid.2017.11.027

76. Yao X, Fan Q, Yao B, Lu P, Rahman SU, Chen D, et al. Codon usage bias analysis of bluetongue virus causing livestock infection. *Front Microbiol* (2020) 11:655. doi: 10.3389/fmicb.2020.00655

77. Zhang Z, Dai W, Wang Y, Lu C, Fan H. Analysis of synonymous codon usage patterns in torque teno sus virus 1 (Ttsuv1). *Arch Virol* (2013) 158:145–54. doi: 10.1007/s00705-012-1480-y

78. Bouquet J, Cherel P, Pavio N. Genetic characterization and codon usage bias of full-length hepatitis e virus sequences shed new lights on genotypic distribution, host restriction and genome evolution. *Infection Genet Evol* (2012) 12(8):1842–53. doi: 10.1016/j.meegid.2012.07.021

79. Roychoudhury S. Mukherjee d. a detailed comparative analysis on the overall codon usage pattern in herpesviruses. *Virus Res* (2010) 148(1-2):31–43. doi: 10.1016/j.virusres.2009.11.018