Check for updates

OPEN ACCESS

EDITED BY Michael Kogut, Agricultural Research Service, United States Department of Agriculture, United States

REVIEWED BY Agusto R. Luzuriaga Neira, American Museum of Natural History, United States Elif Sevim, Ahi Evran University, Türkiye

*CORRESPONDENCE Yeşim Aktürk Dizman ⊠ yesim.akturk@erdogan.edu.tr

RECEIVED 26 January 2025 ACCEPTED 24 June 2025 PUBLISHED 04 July 2025

CITATION

Aktürk Dizman Y (2025) Comprehensive analysis of the codon usage patterns in the polyprotein coding sequences of the honeybee viruses. *Front. Vet. Sci.* 12:1567209. doi: 10.3389/fvets.2025.1567209

COPYRIGHT

© 2025 Aktürk Dizman. 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.

Comprehensive analysis of the codon usage patterns in the polyprotein coding sequences of the honeybee viruses

Yeşim Aktürk Dizman*

Department of Biology, Faculty of Arts and Sciences, Recep Tayyip Erdoğan University, Rize, Türkiye

Honeybee viruses (HVs) are some of the most significant pathogens affecting these insects and are commonly found in beehives across the globe. This viral infection leads to substantial economic losses in the beekeeping industry. To understand the evolution and adaptation of HVs, such as Acute Bee Paralysis Virus (ABPV), Kashmir Bee Virus (KBV), Chronic Bee Paralysis Virus (CBPV), and Sacbrood Virus (SBV), a detailed analysis of codon usage bias (CUB) was conducted, as no prior studies on this topic had been reported. Analysis of nucleotide content and RSCU revealed that the polyprotein coding sequences of the four HVs were rich in A/U nucleotides, with the third base of synonymous codons predominantly A/U. The polyprotein coding sequences showed a higher effective number of codons (ENC) value, suggesting lower CUB. The ENC plot, PR2 plot, and neutrality analyses indicated that natural selection predominantly shapes the codon usage pattern of polyprotein coding sequences, with minimal influence from mutation pressure. Analyses of the codon adaptation index (CAI) and relative codon deoptimization index (RCDI) showed a strong relationship between HVs and their hosts. These findings could offer essential insights into the overall codon usage patterns of HVs and help in understanding the mechanisms that influence codon usage and genetic evolution in HVs.

KEYWORDS

honeybee viruses, polyprotein, codon usage bias, natural selection, host adaptation

1 Introduction

Honeybees are vital pollinators that underpin global biodiversity and agricultural productivity (1, 2). However, over recent decades, honeybee populations have been declining at an alarming rate due to various factors, including climate change, pesticide exposure, habitat loss, and the spread of infectious pathogens (3–5). Among these, honeybee viruses such as Deformed wing virus (DWV), Sacbrood virus (SBV), Black queen cell virus (BQCV), Kashmir bee virus (KBV), Acute bee paralysis virus (ABPV), Israeli acute paralysis virus (IAPV) and Chronic bee paralysis virus (CBPV) play a key role in the health deterioration of honeybee colonies (6–8). SBV and DWV are classified under the genus *Iflavirus* in the family *Iflaviridae*. In contrast, KBV, ABPV, and IAPV belong to the genus *Aparavirus* within the family *Dicistroviridae*, while BQCV is part of the genus *Triatovirus*, also in the family *Dicistroviridae* (9–11). The CBPV remains unclassified (12). These viruses, often vectored by ectoparasitic mites like Varroa destructor, contribute significantly to colony collapse disorder, posing a severe threat to ecosystem stability and agricultural sustainability (5, 13). Understanding the molecular mechanisms driving the evolution and adaptation of these viruses to their hosts is critical for developing effective strategies to safeguard honeybee populations.

10.3389/fvets.2025.1567209

The genetic code is made up of 64 standard codons, with UGA, UAG, and UAA serving as stop codons that signal the end of translation. The remaining 61 codons code for 20 standard amino acids. Of these, tryptophan and methionine are each represented by a single codon, UGG and AUG, respectively, while the other 18 amino acids are coded by two or more codons. The genetic code's degeneracy enables multiple codons, known as synonymous codons, to encode the same amino acid (14, 15). However, these synonymous codons are not utilized with equal frequency, leading to codon usage bias (CUB) (16). CUB is shaped by a variety of factors, including mutational pressure, natural selection, gene length, GC content, and tRNA availability (17, 18). Viral genomes exhibit unique features compared to those of eukaryotes and prokaryotes, including their dependence on hosts for replication, protein synthesis, and transmission. This interaction between viruses and their hosts is believed to play a crucial role in viral survival, adaptation, evasion of the host's immune system, and evolution (19-21). Thus, understanding codon usage in viruses offers insights into molecular evolution, enhances our knowledge of viral gene expression regulation, and supports vaccine development by optimizing the effective expression of viral proteins needed to elicit immunity.

CUB is increasingly recognized as an important tool for understanding virus-host interactions (22, 23). Viruses with codon usage patterns that align closely with those of their hosts may achieve more efficient replication and protein expression, reflecting a process of host adaptation. Conversely, deviations in codon usage may indicate evolutionary constraints or shifts in host-virus dynamics (24, 25). Although CUB has been extensively studied in many genomes (26, 27), research on honeybee viruses remains relatively limited (28–30). Hence, this study seeks to fill this gap by conducting a comparative analysis of codon usage bias in HVs, including ABPV, CBPV, KBV, and SBV, which could ultimately contribute to the development of targeted strategies to manage honeybee health, mitigate the impact of viral pathogens on pollinator populations, and provide a foundation for future research on virus evolution and host adaptation.

2 Materials and methods

2.1 Retrieving coding sequences data

CUB analysis was performed on ABPV, CBPV, KBV, and SBV. Since the codon usage patterns of BQCV, DWV, and IAPV have been previously studied (29), they were not included in the present study. The complete nucleotide sequences of the polyprotein coding sequences from 43 ABPV, 57 CBPV, 8 KBV, and 96 SBV were obtained in FASTA format from the National Center for Biotechnology Information (NCBI).¹ The accession numbers are provided in Supplementary Table 1.

2.2 Codon usage analysis

The total nucleotide content percentages (U, A, C, and G), nucleotide content at the 3rd position of synonymous codons (U3s,

A3s, C3s, and G3s %), as well as the overall GC and AU content percentages, along with their frequencies at the first (GC1%), second (GC2%), and third positions (GC3s %) of synonymous codons, were determined with the CodonW 1.4.2 (31), which provides both menudriven and command-line interfaces to facilitate flexible data analysis.

The effective number of codons (ENC) index measures the extent of bias in synonymous codon usage. The ENC values were computed with CodonW 1.4.2 (31). ENC values range between 20 and 61, with values below 35 suggesting a CUB; where lower values reflect a stronger bias.

2.3 Relative synonymous codon usage analysis

To evaluate synonymous codon usage independently of amino acid composition across various gene samples, the relative synonymous codon usage (RSCU) values for each codon in every sequence were computed using CodonW 1.4.2 (31). An RSCU value above 1.0 demonstrates a codon that is used more frequently (high bias), while a value below 1.0 signifies a codon that is used less frequently (low bias). An RSCU value of 1.0 signifies an absence of bias, indicating equal preference among codons encoding a particular amino acid (32). An RSCU value exceeding 1.6 indicates an overrepresented synonymous codon, whereas a value below 0.6 signifies an underrepresented synonymous codon (33).

2.4 Relative dinucleotide abundance analysis

Relative dinucleotide abundance analysis focuses on analyzing the frequency patterns of dinucleotide pairs within a specific sequence. This analysis offers an understanding of the compositional biases and possible functional roles of dinucleotide patterns in nucleic acid sequences (34). The formula for calculating dinucleotide frequency, as described by Kariin and Burge (35), is $\rho xy = fxy/fxfy$. Here, fx and fy are the frequencies of the individual nucleotides X and Y, fxy represents the observed frequency of the dinucleotide XY, and ρxy denotes the relative frequency of the dinucleotide XY. Dinucleotides with ρxy values exceeding 1.23 were categorized as overrepresented, whereas those below 0.78 were considered underrepresented. The compseq software was used to carry out this analysis,² as it calculates the composition of sequence motifs of a specified length (e.g., dimers, trimers) within the input sequences.

2.5 ENC plot analysis

An ENC plot analysis was carried out to explore the factors affecting CUB in the polyprotein coding sequences of HVs. The ENC plot illustrates the correlation between ENC and GC3 values. If the observed ENC values are close to or align with the expected ENC curve, it indicates that mutation pressure is the main factor shaping codon usage. However,

¹ https://www.ncbi.nlm.nih.gov/genbank/

² https://www.bioinformatics.nl/cgi-bin/emboss/compseq

deviations below the expected curve imply that other factors, such as natural selection, are also influencing the CUB (36). The formula used to compute the expected ENC values was as follows (37):

$$ENC_{exp} = 2 + GC3s + \left(\frac{29}{GC3s^2 + (1 - GC3s)^2}\right)$$

2.6 Neutrality plot analysis

A neutrality plot is commonly used to identify whether mutational forces or evolutionary forces predominantly influence the CUB of a gene. The plot was generated with GC12 displayed on the y-axis and GC3 plotted on the x-axis. In this graph, a regression line slope near 0 indicates the absence of mutation pressure (with natural selection playing a dominant role), while a slope close to 1 signifies complete neutrality (with mutation pressure being dominant) (38).

2.7 Parity rule 2 analysis

Parity rule 2 (PR2) analysis was applied to evaluate the relative roles of natural selection and mutation pressure on the CUB in the polyprotein coding sequences of HVs. In the PR2 plot, the x-axis displays the value of GC bias at the 3rd codon position [G3/ (G3 + C3)], while the y-axis shows the value of AU bias at the 3rd codon position [A3/ (A3 + U3)]. The origin point (0.5, 0.5) signifies an equal balance between A and T (A = T) as well as between G and C (G = C) (39). Departures from the origin point imply the effects of mutation pressure and natural selection on codon usage.

2.8 Codon adaptation index and relative codon deoptimization index analysis

The codon adaptation index (CAI) was determined using the CAI calculation tool provided by the CAIcal server (40), which performs various computations related to codon usage and the adaptation of DNA or RNA sequences to host organisms. CAI analysis is used to estimate gene expression levels and assess how effectively viral genes have adapted to their host organisms by comparing them with the reference host's RSCU. Higher CAI values indicate higher levels of gene expression and greater codon bias; the values range from 0 to 1.

Relative codon deoptimization index (RCDI) analysis was conducted utilizing the RCDI server (41). An RCDI value of 1 signifies that the virus has a codon usage pattern well-adapted to the host. In contrast, a value greater than 1 indicates reduced adaptation (42). An elevated RCDI value signifies a larger deviation from the codon usage pattern of host. The host codon usage was extracted from the codon and codon pair usage tables (CoCoPUTs) (43).

2.9 Correspondence analysis

Correspondence analysis (COA), a multivariate statistical analysis technique, is commonly employed to explore the connections among

samples and variables (44). COA was conducted using the CodonW 1.4.2 (31) with RSCU values for individual codons to investigate the codon usage patterns of the polyprotein coding sequences in ABPV, CBPV, KBV, and SBV. Each polyprotein coding sequences was depicted as a 59-dimensional vector space, representing 59 synonymous codons, excluding TAA, TAG, TGA, ATG, and TGG with each point corresponding to the RSCU values of the synonymous codons. The first two axes were sufficient to explain a larger portion of the data compared to the other axes, so codons were plotted along these two axes.

2.10 General average hydropathicity and aromaticity

The GRAVY index is determined by averaging the hydropathy values of each amino acid. Its score spans from -2 to 2, with a positive value indicating a hydrophobic protein and a negative value indicating a hydrophilic protein (42). AROMA values indicate whether the gene products contain the aromatic amino acids (Tyr, Phe, and Trp) that the codons encode (45). These two indices are employed to examine the influence of natural selection on the codon usage pattern (33).

2.11 Correlation analysis

Correlation analysis was conducted to characterize the relationships between nucleotide contents and the codon usage patterns of the polyprotein coding sequences in ABPV, CBPV, KBV, and SBV. Correlation analyses were conducted using Pearson correlation method with OriginPro 9.0 software, and statistical significance was considered at p < 0.05.

3 Results

3.1 Nucleotide content in the polyprotein coding sequences

In the polyprotein coding sequences of the four HVs, nucleotide A had the highest mean composition at 29.68%, followed by U at 27.46%, C at 21.77%, and G at 21.09% (Table 1). At the 3rd positions of synonymous codons, the nucleotide composition displayed a distinct pattern, with U3s being the most prevalent at 40.01%, followed by A3s at 39.49%, C3s at 22.97%, and G3s at 20.83%. This indicates an enrichment of A/U-ending codons in the polyprotein coding sequences. The average contents of AU and GC were 57.14 and 43.02%, respectively. Furthermore, the GC content varied across different codon positions, with GC1 being the highest at 47.86%, followed by GC12 at 45.31%, GC2 at 42.76%, and GC3s at 34.94%.

3.2 Codon usage bias of the polyprotein coding sequences

The CUB of polyprotein coding sequences in the four HVs was assessed using the effective number of codons (ENC). Generally, a lower ENC value indicates a stronger preference for certain codons.

TABLE 1 Codon usage indices of polyprotein coding sequences in four HVs.

Items	ABPV	CBPV	KBV	SBV	MEAN
A%	34.47	21.40	32.99	29.86	29.68
U%	29.14	25.58	25.77	26.10	26.65
G%	19.29	20.85	19.72	21.77	20.41
C%	17.11	32.17	21.52	14.48	21.32
GC	0.365	0.536	0.410	0.410	0.430
U3s	0.458	0.297	0.382	0.464	0.400
C3s	0.143	0.419	0.236	0.121	0.230
A3s	0.493	0.242	0.466	0.379	0.395
G3s	0.157	0.220	0.159	0.297	0.208
GC3s	0.231	0.539	0.315	0.313	0.349
ENC	44.10	54.10	51.86	49.05	49.78
AU	63.61	46.98	58.75	59.22	57.14
AU3	73.88	42.73	65.50	66.06	62.04
GC1	42.92	53.03	48.05	47.43	47.86
GC2	40.15	48.74	41.19	40.98	42.76
GC12	41.53	50.89	44.62	44.20	45.31
GRAVY	-0.294	0.271	-0.265	-0.251	-0.135
AROMA	0.097	0.078	0.100	0.098	0.093

The average ENC values for the polyprotein coding sequences of ABPV, CBPV, KBV, and SBV were 44.10, 54.10, 51.86, and 49.05, respectively (Figure 1). These values indicate a conserved genomic structure and stable with minimal codon usage bias in all the polyprotein coding sequences analyzed.

3.3 Polyprotein coding sequences preferentially use A- and U-ending codons

We carried out an RSCU analysis to investigate the codon usage patterns in the polyprotein coding sequences of four HVs (Table 2). Out of the 18 preferred codons (RSCU > 1) common to all four HVs, 10 ended in U (UUU, GUU, AUU, UCU, UAU, GCU, CAU, AAU, UGU, and GAU), while 8 ended in A (UUA, CCA, CAA, ACA, AAA, AGA, GAA, and GGA). These results indicate that codons ending in A and U are favored in the polyprotein coding sequences of the four HVs. Almost all the favored and non-favored codons lie in the range from 0.6 to 1.6. Among the 59 codons, 3 were over-represented (RSCU > 1.6) in all the polyprotein coding sequences of the four HVs, namely UUA, AGA, and GGA. Conversely, 7 out of the 59 codons were underrepresented (RSCU < 0.6), such as CUG, AGC, GCG, UAC, UCG, CGG, and GGG. Interestingly, the results also revealed that all overrepresented codons ended with A, while the most underrepresented codons ended with G, suggesting that mutational bias played a key role in influencing the codon usage patterns in polyprotein coding sequences.

To assess if the CUB of the polyprotein coding sequences in the four HVs is affected by their hosts, a comparison was made between the codon usage patterns of the polyprotein coding



sequences and those of the host organisms, including *Apis mellifera*, *Apis cerana*, *Vespa velutina*, and *Varroa destructor*. The results revealed that, for *Apis mellifera*, 37 out of 59 synonymous codons were selected in a similar manner; for *Apis cerana* and *Vespa velutina*, 43 out of 59 synonymous codons were selected identically; and for *Varroa destructor*, 47 out of 59 synonymous codons were selected in the same way. Interestingly, certain codons, including UUU, UUA, AUU, GUU, and UCU, showed similarities between the polyprotein coding sequences and their hosts, suggesting a potential link to virulence in the host species. On the other hand, codons such as UCG, GCG, CAG, AAC, and CGC displayed substantial differences between the polyprotein coding sequences and their hosts. RSCU analysis suggested that compositional constraints, specifically A and U, have largely influenced the preferred codons.

3.4 Impact of relative dinucleotide abundance on codon usage bias

The frequency of dinucleotides impacts codon usage (46). The relative abundances of the 16 dinucleotides in the polyprotein coding sequences of four HVs were ascertained in order to look into the possible influence of dinucleotides on codon usage. As shown in Figure 2, we found that the polyprotein coding sequences did not have uniform distributions of the frequencies of occurrence of each dinucleotide. Specifically, the majority of dinucleotides (AC, AG, AU, GA, GG, and UU) were consistent with the theoretical value, being close to 1, with means of 1.01, 1.00, 1.05, 1.08, 1.03, and 1.03, respectively. Additionally, the results revealed that none of the dinucleotides were over-represented, while CG was underrepresented. Moreover, the RSCU values of the three codons containing CG (UCG, GCG, and CGG) indicate that these codons are not favored. In conclusion, besides the total base composition, the dinucleotide composition and CG suppression were linked to the variation in synonymous codon usage, suggesting that mutational pressure may play a role in limiting codon usage patterns within polyprotein coding sequences.

AA	Codons	RSCU (Mean)	Apis mellifera	Apis cerana	Vespa velutina	Varroa destructor
Phe (F)	<u>UUU</u>	1.25	1.58	1.44	1.39	1.29
	UUC	0.75	0.42	0.56	0.61	0.71
Leu (L)	UUA	1.62	3.07	2.51	1.83	1.40
	UUG	1.52	1.03	1.10	0.90	1.16
	CUU	0.94	0.78	0.96	1.20	1.07
	CUC	0.94	0.25	0.52	1.09	0.63
	CUA	0.61	0.60	0.56	0.61	0.92
	CUG	0.38	0.27	0.35	0.37	0.82
Ile (I)	AUU	1.23	1.28	1.36	1.12	1.15
	AUC	0.74	0.33	0.45	0.55	0.69
	AUA	1.03	1.38	1.19	1.33	1.16
Val (V)	GUU	1.46	1.60	1.48	1.31	1.29
	GUC	0.69	0.41	0.57	0.70	0.78
	GUA	1.12	1.49	1.22	1.25	1.11
	GUG	0.74	0.50	0.72	0.75	0.82
Ser (S)	UCU	1.54	1.57	1.48	2.05	1.17
	UCC	0.95	0.47	0.75	0.70	0.68
	UCA	1.46	1.94	1.42	0.99	1.16
	UCG	0.48	0.58	1.06	1.08	0.90
	AGU	0.99	1.00	0.84	0.75	1.08
	AGC	0.58	0.44	0.45	0.44	1.02
Pro (P)	CCU	1.26	1.12	1.18	1.37	1.21
	CCC	0.60	0.40	0.69	0.75	0.69
	CCA	1.54	1.92	1.36	1.14	1.25
	CCG	0.60	0.56	0.77	0.74	0.86
Thr (T)	ACU	1.20	1.21	1.03	0.93	1.07
	ACC	0.73	0.33	0.51	0.56	0.71
	ACA	1.42	1.98	1.54	1.42	1.33
	ACG	0.65	0.48	0.92	1.10	0.90
Ala (A)	GCU	1.49	1.10	1.02	1.04	1.23
	GCC	0.75	0.43	0.65	0.63	0.80
	GCA	1.26	1.87	1.38	1.34	1.21
	GCG	0.50	0.59	0.95	0.99	0.76
Tyr (Y)	<u>UAU</u>	1.44	1.71	1.60	1.49	1.24
	UAC	0.56	0.29	0.40	0.51	0.76
His (H)	CAU	1.22	1.58	1.36	1.27	1.18
	CAC	0.78	0.42	0.64	0.73	0.82
Gln (Q)	CAA	1.28	1.60	1.52	1.42	1.16
	CAG	0.72	0.40	0.48	0.58	0.84
Asn (N)	AAU	1.26	1.67	1.58	1.40	1.16
	AAC	0.74	0.33	0.42	0.60	0.84
Lys (K)	AAA	1.16	1.70	1.60	1.47	1.32
	AAG	0.84	0.30	0.40	0.53	0.68

TABLE 2 The relative synonymous codon usage frequency of polyprotein coding sequences in HVs and its hosts.

(Continued)

AA	Codons	RSCU (Mean)	Apis mellifera	Apis cerana	Vespa velutina	Varroa destructor
Asp (D)	GAU	1.26	1.63	1.53	1.36	1.15
	GAC	0.74	0.37	0.47	0.64	0.85
Glu (E)	GAA	1.26	1.67	1.48	1.14	1.30
	GAG	0.74	0.33	0.52	0.86	0.70
Cys (C)	UGU	1.16	1.44	1.33	1.34	1.14
	UGC	0.84	0.56	0.67	0.66	0.86
Arg (R)	CGU	0.96	0.69	0.98	1.00	1.09
	CGC	0.93	0.34	0.56	0.48	0.78
	CGA	1.06	1.10	1.41	1.17	1.11
	CGG	0.47	0.24	0.44	0.36	0.66
	AGA	1.72	3.08	1.96	2.32	1.44
	AGG	0.86	0.54	0.64	0.67	0.92
Gly (G)	GGU	0.97	1.20	0.93	1.03	1.11
	GGC	0.71	0.62	0.63	0.61	1.07
	GGA	1.76	1.79	1.74	1.62	1.13
	GGG	0.56	0.39	0.70	0.74	0.70

TABLE 2 (Continued)

The most preferred codon used in coding each amino acid is shown in bold. Blue and green color marked the over-represented (RSCU>1.6) and under-represented (RSCU<0.6) codons, respectively. Codons with underline represent the preferred codons for both HVs and their host.



3.5 Factors influencing codon usage bias in polyprotein coding sequences

To investigate the factors affecting codon usage patterns, ENC plot analysis, PR2 bias analysis, and neutrality analysis were applied to the polyprotein coding sequences. In the ENC plot analysis, some points are close to the theoretical fitting curve, while others fall below or on the curve, suggesting that both natural selection and mutation pressure contribute to the codon usage pattern of the polyprotein coding sequences of four HVs (Figure 3).

In the PR2 analysis, no bias in mutation pressure or selection is present when the plot is centered, with both coordinates at 0.5. The PR2 bias analysis showed that all points are distant from (0.5, 0.5), with the majority of points from the ABPV polyprotein coding sequences falling in the region where A3s/ (A3s + U3s) > 0.5 and G3s/ (G3s + C3s) > 0.5. This indicates a preference for A over U and G over C in the third codon position (Figure 4). Meanwhile, the CBPV polyprotein coding sequences were located in the region where A3s/ (A3s + U3s) < 0.5 and G3s/(G3s + C3s) < 0.5, suggesting that U is preferred over A and C is preferred over G in the third codon position. While the KBV polyprotein coding sequences fell in the region where A3s/ (A3s + U3s) > 0.5 and G3s/ (G3s + C3s) < 0.5, the SBV polyprotein coding sequences were in the region where A3s/ (A3s + U3s) < 0.5 and G3s/ (G3s + C3s) > 0.5. Thus, for KBV polyprotein coding sequences, A is favored over U, and C is favored over G at the third codon position, while for SBV polyprotein coding sequences, U is favored over A, and G is favored over C in the third codon position. These results demonstrate that mutational pressure, combined with factors like natural selection, influences the CUB of polyprotein coding sequences.

A neutrality plot can demonstrate the influence of natural selection and mutation pressure on the nucleotide content of genes. In the neutrality analysis, notable correlations were identified between the indices ($R^2 = 0.0183$, $R^2 = 0.0471$, $R^2 = 0.1568$, and $R^2 = 0.0015$; p < 0.05) for the polyprotein coding sequences of ABPV, CBPV, KBV, and SBV, respectively (Figure 5). The slopes of the regression lines for these polyprotein coding sequences were determined to be 0.1998, 0.2577, 0.2498, and 0.6913, respectively. This indicates that mutation pressure accounted for 19.98, 25.77, 24.98, and 69.13%, respectively, while natural selection contributed 80.02, 74.23, 75.02, and 30.87% to the CUB in the polyprotein coding sequences of ABPV, CBPV, KBV,



FIGURE 3

ENC-plot analysis of the polyprotein coding sequences for ABPV (A), CBPV (B), KBV (C), and SBV (D). ENC values are plotted against the GC3s content, with the solid line representing the expected ENC values for random codon usage based on GC3s.



Parity rule 2 (PR2) plot analysis of the polyprotein coding sequences for ABPV (A), CBPV (B), KBV (C), and SBV (D). The center of the plot (coordinates at 0.5, 0.5) represents a position with no bias.



and SBV, respectively. In summary, natural selection has a more substantial impact than mutation pressure in determining the codon usage patterns of the polyprotein coding sequences of ABPV, CBPV, and KBV. However, mutation pressure is the dominant factor influencing the codon usage bias in the polyprotein coding sequences of SBV.

3.6 Codon usage adaptation in polyprotein coding sequences

The codon adaptation index (CAI) was used to examine how the virus adjusts its codon usage to better adapt to the host. CAI values were calculated by comparing the codon usage of the polyprotein coding sequences to the codon usage of their hosts (Apis mellifera, Apis cerana, Vespa velutina, and Varroa destructor), which served as reference sets (Table 3). Higher CAI values, nearing 1, indicate a stronger adaptation to the host, whereas lower values, closer to 0, reflect weaker adaptation. The mean CAI values for the ABPV polyprotein coding sequences were 0.83, 0.88, 0.89 and 0.67 for the Apis cerana, Apis mellifera, Varroa destructor, and Vespa velutina, respectively. In comparison, the mean CAI values for the CBPV polyprotein coding sequences were 0.64, 0.74, and 0.87 for the Apis cerana, Apis mellifera, Varroa destructor, respectively. The mean CAI value for the KBV polyprotein coding sequences was 0.82 for the Apis mellifera, while the mean CAI values for the SBV polyprotein coding sequences were 0.78 and 0.84 for the Apis cerana, Apis

Virus	Host	CAI	RCDI	
ABPV	Apis cerana 0.83		1.18	
	Apis mellifera	0.88	1.16	
	Varroa destructor	0.89	1.13	
	Vespa velutina	0.67	1.62	
CBPV	Apis cerana	0.64	1.62	
	Apis mellifera	0.74	1.45	
	Varroa destructor	0.87	1.11	
KBV	Apis mellifera	0.82	1.18	
SBV	Apis cerana	0.78	1.11	
	Apis mellifera	0.84	1.14	

TABLE 3 The CAI values and the RCDI values of polyprotein coding sequences in HVs for corresponding hosts.

mellifera, respectively. These values suggest that the polyprotein coding sequences exhibited a high level of host adaptation for all the hosts.

In addition, we conducted a relative codon deoptimization index (RCDI) analysis to evaluate the combined impact of codon biases on gene expression (Table 3). The mean RCDI values for the ABPV polyprotein coding sequences were highest in *Vespa velutina*, followed by *Apis mellifera*, Varroa destructor, and *Apis cerana*. For the CBPV polyprotein coding sequences, the highest mean RCDI values were observed in *Apis cerana*, followed by *Apis mellifera* and *Varroa*

destructor. In the case of the SBV polyprotein coding sequences, the highest mean RCDI values were found in *Apis mellifera*, followed by *Apis cerana*. The findings indicated that codon usage deoptimization of ABPV, CBPV, and SBV was the highest for *Vespa velutina*, *Apis cerana* and *Apis mellifera*, respectively. The results showed that codon usage deoptimization was highest for *Vespa velutina* in ABPV, for *Apis cerana* in CBPV, and for *Apis mellifera* in SBV.

3.7 Correspondence analysis of codon bias in polyprotein coding sequences

COA was conducted to evaluate variations in synonymous codon usage based on the RSCU values within the polyprotein coding sequences of four HVs (ABPV, CBPV, KBV, and SBV). For ABPV, the first two axes accounted for 37.50 and 19.76% of the variation, respectively, while for CBPV, these values were 76.44 and 6.26%. Similarly, the first two axes for KBV explained 41.95 and 23.71% of the variation, and for SBV, they accounted for 32.88 and 12.52%, respectively. This indicates that codon usage bias is primarily driven by the values of the first and second axes. When codons were ranked based on RSCU values along these two primary axes, a clear separation was observed: A- and U-ending codons were distributed along axis 1, while C- and G-ending codons were dispersed along axis 2 (Figure 6). This pattern highlights the distinct distribution of codons with different base endings across the two axes.

3.8 Correlation analysis between CUB indices

The analysis of CUB in polyprotein coding sequences revealed significant correlations between various indices (Table 4). Specifically, while GC3s, C3s, and G3s demonstrated a high negative correlation with CAI, U3s and A3s demonstrated a robust positive correlation. The results suggest that codon usage in polyprotein coding sequences is associated with gene expression. The selection of specific codons in the polyprotein coding sequences is not arbitrary; instead, it appears to be related to the level of gene expression. A3s and U3s show a positive correlation with CAI, which implies that A/U-rich codons are favored in genes with high expression, while G3s and C3s exhibit a negative correlation, indicating that G/C-rich codons are linked to lower expression levels. This implies that there is selective pressure for A/U-rich codons to enhance translation efficiency in these polyprotein coding sequences. Moreover, we performed a correlation analysis between ENC and CAI, which revealed no significant correlation between the two factors in the polyprotein coding sequences.

Additionally, correlation analysis was conducted to assess the relationship between codon usage bias and the GRAVY and AROMA



Frontiers in Veterinary Science

TABLE 4 Correlation analysis of codon usage indices in the polyprotein coding sequences of HVs.

Indices	CAI	GC	GC12	Axis 1	Axis 2
U3s	0,94,658*	0	0	0	0
C3s	-0,98,178*	0	0	0	0
A3s	0,95,235*	0	0	0	0
G3s	-0,94,617*	0	0	0	0
GC3s	-0,96,622*	0,99,995*	0,99,799*	0	0
GC12	0	0,99,738*	1	0	0
ENC	0	0	0	0	0
GRAVY	-0,99,502*	0,97,389*	0,95,694*	0	0
AROMA	0,982*	0	0	0	0

**p* value < 0.05.

scores (Table 4). GRAVY exhibited a significant positive correlation with GC3s, GC12, and GC, but a significant negative correlation with A and CAI. However, AROMA is only positively correlated with CAI. Neither AROMA nor GRAVY show any correlation with Axis 1 and Axis 2. It was found that the aromaticity and hydrophobicity of the polyprotein coding sequences were weakly associated with the CUB, highlighting the role of natural selection in shaping the codon usage pattern of these sequences.

4 Discussion

Viruses pose a serious threat to the health and survival of honeybees (6, 47). Gaining insight into the evolution and host adaptation mechanisms of viruses that infect honeybees is essential for devising effective and environmentally friendly strategies to manage these diseases. Thus, examining CUB is crucial for investigating genetic evolution and comprehending the characteristics of gene expression (48). CUB is commonly observed across viruses, bacteria, animals, fungi, and plants (33, 49-52). The 'mutationselection drift' theory has been utilized to describe the origin of CUB in genes (53, 54). This theory suggests that evolutionary factors, including the mutation pressure, selection of compositional constraints, and genetic drift within a population, may significantly influence CUB (55). To date, the codon usage patterns of the polyprotein coding sequences in four HVs (ABPV, CBPV, KBV, and SBV) have not been thoroughly explored. In this research, we conducted a systematic analysis of the codon usage patterns and the factors influencing CUB in the polyprotein coding sequences of these four viruses.

Nucleotide composition is a key factor in shaping codon usage in both genes and genomes. The nucleotide content analysis demonstrated that A was the most prevalent in the polyprotein coding sequences of the four HVs. Additionally, the third codon position of these sequences was rich in A/U and poor in G/C, a pattern that aligns with findings in black queen cell virus (BQCV) (29), deformed wing virus (DWV) (29), Israeli acute paralysis virus (IAPV) (29), transmissible gastroenteritis virus (TGEV) (56), duck hepatitis virus 1 (DHV-1) (57), and invertebrate iridescent virus 6 (IIV6) (58). The significant bias toward A and U nucleotides, along with a strong preference for A- and U-ending codons in the polyprotein coding sequences of the four HVs, led us to evaluate the overall CUB in these sequences using ENC analysis. The analysis showed a slight bias and relatively stable codon usage in the polyprotein coding sequences. Our findings align with those of RNA viruses, which typically exhibit a low CUB (29, 59, 60). Earlier studies proposed that the use of a diverse set of codons to encode amino acids (low CUB) and reduced gene expression in RNA viruses may minimize competition for the translation mechanism between the host and the virus, thereby enhancing the virus's replication rate in the host genome (61, 62).

RSCU is a crucial metric for analyzing the codon usage bias of species. Based on RSCU analysis, A/U-ended codons were the most preferred and over-represented in the polyprotein coding sequences of the four HVs, while G/C-ended codons were the most under-represented. Organisms with AT-rich genomes, such as Taenia saginata, Hemerocallis citrina, and Polygonatum species, typically favor A or T in the third position of their coding sequences (63-65). In contrast, GC-rich species, like bacteria and fungi, tend to prefer G or C in the same position (66, 67). As a result, the unique compositional restrictions (specifically A and U) may be responsible for the CUB observed in the polyprotein coding sequences of these four HVs. This result aligns with previous study (29, 30, 68) but contrasts with another study (69), which found that C3s and G3s were more commonly used than U3s and A3s. Furthermore, we compared the host's RSCU values with those of the polyprotein coding sequences. The results indicated that the codon usage pattern of the polyprotein coding sequences closely resembles that of the host. It has been noted that the virus can translate more effectively when the same synonymous codon is used (70).

Moreover, the analysis of relative dinucleotide abundance showed a distinct bias in the usage of dinucleotides within the four HVs polyprotein coding sequences, with a pronounced underrepresentation of CG dinucleotides. This observation is consistent with the findings that many RNA viruses decrease CG dinucleotides (56), which is thought to increase the stability and effectiveness of viral mRNA translation (71). The biased usage patterns of dinucleotides in HVs polyprotein coding sequences may play a role in its host adaptation and evolutionary dynamics mechanisms, highlighting the need for further research into the functional implications of these patterns on viral pathogenicity and fitness. In general, RNA viruses adjust to changes in their host and environment by modifying the structure of their genomes (72).

Codon usage bias, a crucial factor in virus evolution, is affected by multiple elements, such as mutational pressure, natural selection, the content of the genomic region, and the length of the gene (73). Earlier research on enterovirus A and Venezuelan equine encephalitis virus has proposed that the CUB in these viruses is influenced by natural selection, as evidenced by nucleotide content comparisons, ENC plot, PR2 plot and neutrality plot analyses (62, 74). To evaluate whether both mutational pressure and natural selection have impacted viral codon usage patterns, we conducted ENC plot and PR2 plot analyses. The findings suggested that mutational pressure may not be the sole factor influencing codon usage patterns, with natural selection also likely contributing to the codon usage patterns of the four HVs polyprotein coding sequences, in line with previous research (75, 76). Neutrality analysis was also conducted to evaluate the relative contributions of natural selection and mutational pressure in shaping the CUB. The findings revealed that natural selection was the dominant factor affecting the CUB in the four HVs polyprotein coding sequences, consistent with results from previous studies (34, 77). Furthermore, strong correlations were observed between GRAVY/AROMA and nucleotide content. AROMA and GRAVY exhibited a slight correlation with CAI, indicating that the properties of viral proteins also play a role in the variation of codon usage in HVs' polyprotein coding sequences, highlighting the influence of natural selection in shaping the codon usage bias of these sequences.

In addition, COA analysis of RSCU showed that the first axis accounted for only part of the diversity in codon usage. Hence, we determined that natural selection, along with multiple other factors, probably influences the selective restrictions on codon bias in the polyprotein coding sequences of HVs. This finding is consistent with previous research (78, 79).

Host-parasite interactions have a major impact on the evolution and dynamics of infectious diseases (80). Various research have demonstrated that codon usage patterns are vital in virus-host interactions (81, 82). The CAI is commonly employed as a metric for gene expression and to evaluate how viral genes adapt to their hosts (83). In our study, CAI analysis revealed that the HVs polyprotein coding sequences demonstrated a high degree of host adaptation across all hosts. These four HVs are capable of replicating efficiently in multiple hosts. As a result, we suggest that these HVs likely maintain a dynamic balance between codon adaptation and codon deoptimization, enabling effective replication cycles across hosts with diverse codon usage patterns. The lack of a significant correlation between CAI and ENC in the HVs polyprotein coding sequences indicates that codon adaptation in these viruses is influenced by factors beyond mutational bias or natural selection for uniform codon usage (62, 84). In the present study, a positive correlation was observed between A3s and U3s and CAI, indicating a preference for A/U-rich codons in highly expressed genes (55, 85). Conversely, G3s and C3s showed a negative correlation with CAI, suggesting that G/C-rich codons are associated with lower gene expression levels (86, 87). These patterns imply that there is selective pressure favoring A/U-rich codons to enhance translational efficiency, particularly in polyprotein-coding genes. This observation aligns with previous findings in various viral genomes, where codon usage bias has been shown to reflect host-driven selection for efficient translation. For example, studies on RNA viruses have demonstrated a tendency toward A/U-rich codon usage in highly expressed genes, likely driven by host tRNA abundance and translational selection (88, 89).

In conclusion, this study effectively demonstrates that the intricate codon usage patterns in the polyprotein coding sequences of four HVs are influenced by a complex interaction of factors, including natural selection, mutation pressure, and nucleotide compositional constraints. The CUB in polyprotein coding sequences of HVs was found to be low, predominantly driven by natural selection. Additionally, factors like dinucleotide frequencies, aromaticity, and hydrophobicity also play a role in shaping the codon usage pattern. The codon usage patterns in the polyprotein coding sequences of HVs were also observed to be affected by their host. In summary, our study deepens the understanding of the evolution of HVs polyprotein coding sequences, host-virus interactions, and the molecular mechanisms that drive viral adaptation.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

YA: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study has been supported by the Recep Tayyip Erdoğan University Development Foundation (Grant no. 02025001021248).

Conflict of interest

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

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

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.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2025.1567209/ full#supplementary-material

10.3389/fvets.2025.1567209

References

1. Mojgani N, Bagheri M, Ashique S, Islam A, Moharrami M, Modirrousta H, et al. Honeybee defense mechanisms: role of honeybee gut microbiota and antimicrobial peptides in maintaining colony health and preventing diseases. *Microb Pathog.* (2025) 198:107161. doi: 10.1016/j.micpath.2024.107161

2. Gallai N, Salles JM, Settele J, Vaissière BE. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecol Econ.* (2009) 68:810–21. doi: 10.1016/j.ecolecon.2008.06.014

3. Staveley JP, Law SA, Fairbrother A, Menzie CA. A causal analysis of observed declines in managed honey bees (*Apis mellifera*). *Hum Ecol Risk Assess.* (2014) 20:566–91. doi: 10.1080/10807039.2013.831263

4. Brutscher LM, McMenamin AJ, Flenniken ML. The buzz about honey bee viruses. *PLoS Pathog.* (2016) 12:e1005757–7. doi: 10.1371/journal.ppat.1005757

5. Traynor KS, Mondet F, de Miranda JR, Techer M, Kowallik V, Oddie MAY, et al. *Varroa destructor*: a complex parasite, crippling honey bees worldwide. *Trends Parasitol.* (2020) 36:592–606. doi: 10.1016/j.pt.2020.04.004

6. Chen YP, Siede R. Honey bee viruses. Adv Virus Res. (2007) 70:33-80. doi: 10.1016/S0065-3527(07)70002-7

7. Salkova D, Balkanska R, Shumkova R, Lazarova S, Radoslavov G, Hristov P. Molecular detection and phylogenetic relationships of honey bee-associated viruses in bee products. *Vet Sci.* (2024) 11:369. doi: 10.3390/vetsci11080369

8. Shumkova R, Neov B, Sirakova D, Georgieva A, Gadjev D, Teofanova D, et al. Molecular detection and phylogenetic assessment of six honeybee viruses in *Apis mellifera* L. colonies in Bulgaria. *PeerJ*. (2018) 6:1–16. doi: 10.7717/peerj.5077

 Procházková M, Škubník K, Füzik T, Mukhamedova L, Přidal A, Plevka P. Virion structures and genome delivery of honeybee viruses. *Curr Opin Virol.* (2020) 45:17–24. doi: 10.1016/j.coviro.2020.06.007

10. Valles SM, Chen Y, Firth AE, Guérin DMA, Hashimoto Y, Herrero S, et al. ICTV virus taxonomy profile: Iflaviridae. *J Gen Virol.* (2017) 98:527–8. doi: 10.1099/jgv.0.000757

11. Valles SM, Chen Y, Firth AE, Guérin DMA, Hashimoto Y, Herrero S, et al. ICTV virus taxonomy profile: Dicistroviridae. *J Gen Virol.* (2017) 98:355–6. doi: 10.1099/jgv.0.000756

12. Zavatta L, Bortolotti L, Catelan D, Granato A, Guerra I, Medrzycki P, et al. Spatiotemporal evolution of the distribution of chronic bee paralysis virus (CBPV) in honey bee colonies. *Virology*. (2024) 598:110191. doi: 10.1016/j.virol.2024.110191

13. Noël A, Le Conte Y, Mondet F. Varroa destructor: how does it harm Apis mellifera honey bees and what can be done about it? Emerg Top Life Sci. (2020) 4:45–57. doi: 10.1042/ETLS20190125

14. Iriarte A, Lamolle G, Musto H. Codon usage bias: an endless tale. J Mol Evol. (2021) 89:589–93. doi: 10.1007/s00239-021-10027-z

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

16. Grantham R, Gautier C, Gouy M, Mercier R, Pavé A. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* (1980) 8:197–r62. doi: 10.1093/nar/8.1.197-c

17. Shah P, Gilchrist MA. Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc Natl Acad Sci U S A*. (2011) 108:10231–6. doi: 10.1073/pnas.1016719108

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

19. Day T, Kennedy DA, Read AF, Gandon S. Pathogen evolution during vaccination campaigns. *PLoS Biol.* (2022) 20:e3001804–21. doi: 10.1371/journal.pbio.3001804

20. Rahman SU, Rehman HU, Rahman IU, Rauf A, Alshammari A, Alharbi M, et al. Analysis of codon usage bias of lumpy skin disease virus causing livestock infection. *Front Vet Sci.* (2022) 9:1071097. doi: 10.3389/fvets.2022.1071097

21. Zhou J, Gao Z, Zhang J, Ding Y, Stipkovits L, Szathma S, et al. The analysis of codon bias of foot-and-mouth disease virus and the adaptation of this virus to the hosts. *Infect Genet Evol.* (2013) 14:105–10. doi: 10.1016/j.meegid.2012.09.020

22. Simón D, Cristina J, Musto H. Nucleotide composition and codon usage across viruses and their respective hosts. *Front Microbiol.* (2021) 12:1–9. doi: 10.3389/fmicb.2021.646300

23. He Z, Qin L, Xu X, Ding S. Evolution and host adaptability of plant RNA viruses: research insights on compositional biases. *Comput Struct Biotechnol J.* (2022) 20:2600–10. doi: 10.1016/j.csbj.2022.05.021

24. Kaushik R, Kumar N, Yadav P, Sircar S, Shete-Aich A, Singh A, et al. Comprehensive genomics investigation of neboviruses reveals distinct codon usage patterns and host specificity. *Microorganisms.* (2024) 12:696. doi: 10.3390/microorganisms12040696

25. Bahir I, Fromer M, Prat Y, Linial M. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol.* (2009) 5:1–14. doi: 10.1038/msb.2009.71

26. Zhao Y, Zhang S. Comparative analysis of codon usage bias in six Eimeria genomes. Int J Mol Sci. (2024) 25:1–16. doi: 10.3390/ijms25158398

27. Luo XL, Liu Q, Xiong Y, Ye C, Jin D, Xu J. Genome-wide analysis of synonymous codon usage in Huaiyangshan virus and other bunyaviruses. *J Basic Microbiol.* (2015) 55:1374–83. doi: 10.1002/jobm.201500233

28. Shi SL, Xia RX. Codon usage in the Iflaviridae family is not diverse though the family members are isolated from diverse host taxa. *Viruses.* (2019) 11:1087. doi: 10.3390/v11121087

29. Yoon KA, Jin Kim W, Hyeock LS. Comparative analysis of codon usage patterns of venom gland-specific honeybee-infecting viruses in solitary hunting and social wasps. *J Asia Pac Entomol.* (2024) 27:102276. doi: 10.1016/j.aspen.2024.102276

30. Chantawannakul P, Cutler RW. Convergent host-parasite codon usage between honeybee and bee associated viral genomes. *J Invertebr Pathol.* (2008) 98:206–10. doi: 10.1016/j.jip.2008.02.016

31. Peden JF. Analysis of codon usage. PhD dissertation. Nottingham: University of Nottingham (1999).

32. Beelagi MS. Synonymous codon usage pattern among the S, M, and L segments in crimean-Congo hemorrhagic fever causing virus. *Bioinformation*. (2021) 17:479–91. doi: 10.6026/97320630017479

33. Xu Q, Chen H, Sun W, Zhu D, Zhang Y, Chen JL, et al. Genome-wide analysis of the synonymous codon usage pattern of *Streptococcus suis*. *Microb Pathog*. (2021) 150:104732. doi: 10.1016/j.micpath.2021.104732

34. Tan X, Bao S, Lu X, Lu B, Shen W, Jiang C. Comprehensive analysis of codon usage bias in human papillomavirus type 51. *Pol J Microbiol.* (2024) 73:455–65. doi: 10.33073/pjm-2024-036

35. Kariin S, Burge C. Dinucleotide relative abundance extremes: a genomic signature. *Trends Genet.* (1995) 11:283–90. doi: 10.1016/S0168-9525(00)89076-9

36. Yang M, Liu J, Yang W, Li Z, Hai Y, Duan B, et al. Analysis of codon usage patterns in 48 Aconitum species. *BMC Genomics*. (2023) 24:703–13. doi: 10.1186/s12864-023-09650-5

37. Wright F. The "effective number of codons" used in a gene. Gene. (1990) 87:23-9. doi: 10.1016/0378-1119(90)90491-9

38. Wang L, Zhao H, Wang Z, Ding S, Qin L, Jiang R, et al. An evolutionary perspective of codon usage pattern, dinucleotide composition and codon pair bias in prunus necrotic ringspot virus. *Genes (Basel).* (2023) 14:1712. doi: 10.3390/genes14091712

39. Sueoka N. Near homogeneity of PR2-bias fingerprints in the human genome and their implications in phylogenetic analyses. *J Mol Evol.* (2001) 53:469–76. doi: 10.1007/s002390010237

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

41. Puigbò P, Aragonès L, Garcia-Vallvé S. RCDI/eRCDI: a web-server to estimate codon usage deoptimization. *BMC Res Notes*. (2010) 3:2–5. doi: 10.1186/1756-0500-3-87

42. He Z, Gan H, Liang X. Analysis of synonymous codon usage bias in potato virus M and its adaption to hosts. Viruses. (2019) 11:752. doi: 10.3390/v11080752

43. Alexaki A, Kames J, Holcomb DD, Athey J, Santana-Quintero LV, Lam PVN, et al. Codon and codon-pair usage tables (CoCoPUTs): facilitating genetic variation analyses and recombinant gene design. *J Mol Biol.* (2019) 431:2434–41. doi: 10.1016/j.jmb.2019.04.021

44. Xiong B, Wang T, Huang S, Liao L, Wang X, Deng H, et al. Analysis of codon usage bias in xyloglucan endotransglycosylase (XET) genes. *Int J Mol Sci.* (2023) 24:6108. doi: 10.3390/ijms24076108

45. Wei L, He J, Jia X, Qi Q, Liang Z, Zheng H, et al. Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. *BMC Evol Biol.* (2014) 14:1–12. doi: 10.1186/s12862-014-0262-4

46. Li B, Wu H, Miao Z, Hu L, Zhou L, Lu Y. Codon usage of hepatitis E viruses: a comprehensive analysis. *Front Microbiol.* (2022) 13:938651. doi: 10.3389/fmicb.2022.938651

47. Tantillo G, Bottaro M, Di Pinto A, Martella V, Di Pinto P, Terio V. Virus infections of honeybees *Apis mellifera*. *Ital J Food Saf.* (2015) 4:157–68. doi: 10.4081/ijfs.2015.5364

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

49. Yang X, Wang Y, Gong W, Li Y. Comparative analysis of the codon usage pattern in the chloroplast genomes of *Gnetales* species. *Int J Mol Sci.* (2024) 25:10622. doi: 10.3390/ijms251910622

50. Sun J, Zhao W, Wang R, Zhang W, Li G, Lu M, et al. Analysis of the codon usage pattern of HA and NA genes of H7N9 influenza a virus. *Int J Mol Sci.* (2020) 21:7129. doi: 10.3390/ijms21197129

51. Montaña-Lozano P, Balaguera-Reina SA, Prada-Quiroga CF. Comparative analysis of codon usage of mitochondrial genomes provides evolutionary insights into reptiles. *Gene.* (2023) 851:146999. doi: 10.1016/j.gene.2022.146999

52. Li Q, Luo Y, Sha A, Xiao W, Xiong Z, Chen X, et al. Analysis of synonymous codon usage patterns in mitochondrial genomes of nine amanita species. *Front Microbiol.* (2023) 14:1134228. doi: 10.3389/fmicb.2023.1134228

53. Bulmer M. The selection-mutation-drift theory of synonymous codon usage. *Genetics.* (1991) 129:897–907. doi: 10.1093/genetics/129.3.897 54. Hershberg R, Petrov DA. Selection on codon bias. Annu Rev Genet. (2008) 42:287–99. doi: 10.1146/annurev.genet.42.110807.091442

55. Sinha K, Jana S, Pramanik P, Bera B. Selection on synonymous codon usage in soybean (*Glycine max*) WRKY genes. *Sci Rep.* (2024) 14:26530. doi: 10.1038/s41598-024-77156-5

56. Cheng S, Wu H, Chen Z. Evolution of transmissible gastroenteritis virus (TGEV): a codon usage perspective. *Int J Mol Sci.* (2020) 21:1–17. doi: 10.3390/ijms212 17898

57. Zhao Y, Su X, He D. Codon usage analysis reveals distinct evolutionary patterns and host adaptation strategies in duck hepatitis virus 1 (DHV-1) phylogroups. *Viruses.* (2024) 16:1380. doi: 10.3390/v16091380

58. Aktürk Dizman Y. Analysis of codon usage bias of exonuclease genes in invertebrate iridescent viruses. *Virology.* (2024) 593:110030. doi: 10.1016/j.virol.2024.110030

59. Kim H, Cho M, Son HS. Comparative analysis of codon usage patterns in rift valley fever virus. *Genet Mol Biol.* (2020) 43:e20190240–8. doi: 10.1590/1678-4685-GMB-2019-0240

60. Tort FL, Castells M, Cristina J. A comprehensive analysis of genome composition and codon usage patterns of emerging coronaviruses. *Virus Res.* (2020) 283:197976–7. doi: 10.1016/j.virusres.2020.197976

61. Roy A, Guo F, Singh B, Gupta S, Paul K, Chen X, et al. Base composition and host adaptation of the SARS-CoV-2: insight from the codon usage perspective. *Front Microbiol.* (2021) 12:1–11. doi: 10.3389/fmicb.2021.548275

62. Wang H, Liu S, Lv Y, Wei W. Codon usage bias of Venezuelan equine encephalitis virus and its host adaption. *Virus Res.* (2023) 328:199081. doi: 10.1016/j.virusres.2023.199081

63. Zhang K, Wang Y, Zhang Y, Shan X. Codon usage characterization and phylogenetic analysis of the mitochondrial genome in *Hemerocallis citrina*. *BMC Genomic Data*. (2024) 25:6–15. doi: 10.1186/s12863-024-01191-4

64. Shi N, Yuan Y, Huang R, Wen G. Analysis of codon usage patterns in complete plastomes of four medicinal *Polygonatum* species (Asparagaceae). *Front Genet*. (2024) 15:1–14. doi: 10.3389/fgene.2024.1401013

65. Rahman SU, Rehman HU, Rahman IU, Khan MA, Rahim F, Ali H, et al. Evolution of codon usage in *Taenia saginata* genomes and its impact on the host. *Front Vet Sci.* (2023) 9:1021440. doi: 10.3389/fvets.2022.1021440

66. Bansal S, Mallikarjuna MG, Balamurugan A, Nayaka SC, Prakash G. Composition and codon usage pattern results in divergence of the zinc binuclear cluster (Zn(II)2Cys6) sequences among Ascomycetes plant pathogenic fungi. *J Fungi*. (2022) 8:1134. doi: 10.3390/jof8111134

67. Soman S, Chattopadhyay S, Ram S, Nandicoori VK, Arimbasseri GA. Codon optimality has minimal effect on determining translation efficiency in *Mycobacterium tuberculosis. Sci Rep.* (2023) 13:1–11. doi: 10.1038/s41598-022-27164-0

68. Deb B, Uddin A, Chakraborty S. Codon usage pattern and its influencing factors in different genomes of hepadnaviruses. *Arch Virol.* (2020) 165:557–70. doi: 10.1007/s00705-020-04533-6

69. Zhong J, Li Y, Zhao S, Liu S, Zhang Z. Mutation pressure shapes codon usage in the GC-rich genome of foot-and-mouth disease virus. *Virus Genes*. (2007) 35:767–76. doi: 10.1007/s11262-007-0159-z

70. Tao J, Yao H. Comprehensive analysis of the codon usage patterns of polyprotein of Zika virus. *Prog Biophys Mol Biol.* (2020) 150:43–9. doi: 10.1016/j.pbiomolbio.2019.05.001

71. Upadhyay M, Vivekanandan P. Depletion of CpG dinucleotides in papillomaviruses and polyomaviruses: a role for divergent evolutionary pressures. *PLoS One.* (2015) 10:1–16. doi: 10.1371/journal.pone.0142368

72. LaTourrette K, Garcia-Ruiz H. Determinants of virus variation, evolution, and host adaptation. *Pathogens*. (2022) 11:1039. doi: 10.3390/pathogens11091039

73. Jenkins GM, Holmes EC. The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res.* (2003) 92:1–7. doi: 10.1016/S0168-1702(02)00309-X

74. Zeng L, Chen M, Wang M, Zhu L, Yan J, Zhang X, et al. Enterovirus a shows unique patterns of codon usage bias in conventional versus unconventional clade. *Front Cell Infect Microbiol.* (2022) 12:1–13. doi: 10.3389/fcimb.2022.941325

75. Liu Z, Zhao S, Jin X, Wen X, Ran X. Host and structure-specific codon usage of G genotype (VP7) among group a rotaviruses. *Front Vet Sci.* (2024) 11:1–11. doi: 10.3389/fvets.2024.1438243

76. Guo F, Tan H, Yang J, Jia R, Wang R, Wu L, et al. Insight into the codon usage patterns and adaptation of Tembusu virus. *Poult Sci.* (2025) 104:104651. doi: 10.1016/j.psj.2024.104651

77. Aktürk Dizman Y. Codon usage bias in the photolyase gene of Amsacta moorei entomopoxvirus. *Biol Bull.* (2024) 51:S52–65. doi: 10.1134/S1062359024610164

78. Mazumder TH, Alqahtani AM, Alqahtani T, Emran TB, Aldahish AA, Uddin A. Analysis of codon usage of speech gene foxp2 among animals. *Biology (Basel)*. (2021) 10:1–15. doi: 10.3390/biology10111078

79. Aktürk Dizman Y. A comprehensive comparative analysis on the codon usage bias of DNA polymerase genes in invertebrate iridescent viruses. *Recep Tayyip Erdogan Univ J Sci Eng.* (2024) 5:32–48. doi: 10.53501/rteufemud.1416072

80. Penczykowski RM, Laine AL, Koskella B. Understanding the ecology and evolution of host-parasite interactions across scales. *Evol Appl*. (2016) 9:37–52. doi: 10.1111/eva.12294

81. Ma MR, Ha XQ, Ling H, Wang ML, Zhang FX, Di Zhang S, et al. The characteristics of the synonymous codon usage in hepatitis B virus and the effects of host on the virus in codon usage pattern. *Virol J.* (2011) 8:1–10. doi: 10.1186/1743-422X-8-544

82. Das JK, Chakraborty S, Roy S. A scheme for inferring viral-host associations based on codon usage patterns identifies the most affected signaling pathways during COVID-19. *J Biomed Inform*. (2021) 118:103801. doi: 10.1016/j.jbi.2021.103801

83. Wang X, Sun J, Lu L, Pu FY, Zhang DR, Xie FQ. Evolutionary dynamics of codon usages for peste des petits ruminants virus. *Front Vet Sci.* (2022) 9:968034. doi: 10.3389/fvets.2022.968034

84. He Z, Dong Z, Qin L, Gan H. Phylodynamics and codon usage pattern analysis of broad bean wilt virus 2. *Viruses.* (2021) 13:198. doi: 10.3390/v13020198

85. Benisty H, Hernandez-Alias X, Weber M, Anglada-Girotto M, Mantica F, Radusky L, et al. Genes enriched in a/T-ending codons are co-regulated and conserved across mammals. *Cell Syst.* (2023) 14:312–323.e3. doi: 10.1016/j.cels.2023.02.002

86. Newman ZR, Young JM, Ingolia NT, Barton GM. Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. *Proc Natl Acad Sci USA*. (2016) 113:E1362–71. doi: 10.1073/pnas.1518976113

87. Gao Y, Lu Y, Song Y, Jing L. Analysis of codon usage bias of WRKY transcription factors in *Helianthus annuus. BMC Genomic Data*. (2022) 23:46–12. doi: 10.1186/s12863-022-01064-8

88. Qi X, Wei C, Li Y, Wu Y, Xu H, Guo R, et al. The characteristic of the synonymous codon usage and phylogenetic analysis of hepatitis B virus. *Genes Genom.* (2020) 42:805–15. doi: 10.1007/s13258-020-00932-w

89. Shackelton LA, Parrish CR, Holmes EC. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J Mol Evol.* (2006) 62:551–63. doi: 10.1007/s00239-005-0221-1