International Journal Of Engineering Research And Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 21, Issue 5 (May 2025), PP 22-33

A Study on the Human Adaptation of the IMV Membrane Protein Gene in Monkeypox Virus

Jiarui Kang, Gun Li
School of Electronic Information Engineering, Xi'An Technological University, Xi'An, Shaanxi, China

Abstract: *This study aims to comprehensively understand the evolutionary characteristics and host adaptability of Monkeypox virus through systematic analysis. Initially, we computed key metrics such as nucleotide composition, Effective Number of Codons (ENC), Codon Adaptation Index (CAI), and Relative Synonymous Codon Usage (RSCU). Subsequently, we employed ENC-plot, PR2-plot, and neutrality plot analyses to further explore the impact of natural selection and mutations on virus evolution. Finally, a comprehensive analysis of codon usage patterns in the membrane protein gene of Monkeypox virus encompassed 11 gene sequences.This revision aims to streamline and enhance the clarity and precision of your research paper while maintaining its scholarly rigor and scientific impact. Our study reveals significant similarities in nucleotide composition among these 11 genes, all of which preferentially use codons ending in adenine (A) or thymine (T). These genes exhibit high ENC values and low CAI values. Furthermore, comparative analysis of RSCU between the IMV membrane protein gene and its host indicates that codons such as CCA, AGA, GGA, TTA, GAA, CCC, ACT, and ACG are significantly overrepresented in the gene sequences compared to the host, whereas codons like CTG, GTG, AGC, GCC, AGG, ATC, and ACC are notably underrepresented. ENC-plot, neutrality plot, and PR2-plot analyses demonstrate that the codon preference in the Monkeypox virus IMV membrane protein gene is predominantly influenced by mutations. The codon usage pattern of the monkeypox virus IMV membrane protein gene suggests a relatively low expression level within its host species. It exhibits a preference for codons ending with adenine (A) or thymine (T), indicating a weak bias in codon usage that tends to mirror its host's preferences. Beyond mutation pressures, natural selection also influences codon bias.*

Keywords: *Monkeypox virus, IMV membrane protein gene, codon usage preference, nucleotide mutations*

Date of Submission: 01-05-2025 Date of acceptance: 08-05-2025

I. Introduction

Monkeypox is a zoonotic disease[1] caused by the Orthopoxvirus genus[2], which also includes Variola Virus (VARV), Cowpox Virus (CPXV), and Vaccinia Virus (VACV)[3]. Transmission primarily occurs through direct contact with the bodily fluids of infected individuals, such as saliva, respiratory secretions, or exudates from skin lesions[4].The monkeypox virus has a double-stranded DNA genome approximately 197 kb in length, encoding around 190 genes[5]. It exists in two infectious forms: intracellular mature virus (IMV) and extracellular enveloped virus (EEV)[6]. The first documented human case of monkeypox was reported in 1970, involving a nine-month-old infant in the Democratic Republic of Congo[7]. Initially, all reported cases were localized endemic infections within African countries. However, during the 1996-1997 outbreak, human-to-human transmission intensified significantly[8], signaling an escalating public health concern.Since 2022, a global resurgence of monkeypox has been observed, with a dramatic increase in infections. In response, the World Health Organization (WHO) officially declared the "2022–2023 Monkeypox Outbreak" a public health emergency[9].

Research on viral codon usage has yielded significant applications across various fields. For example, Celina et al[10] discovered that the codon usage preferences of Crimean-Congo hemorrhagic fever virus (CCHFV) differ among various tick hosts, a finding that is closely associated with the evolutionary and genetic adaptation of the virus to its hosts. In addition, Kaushik et al[11] conducted a systematic study on the codon usage patterns of Neboviruses (NeVs), revealing marked differences across different host species.Despite these advances, the application of codon usage analysis to monkeypox virus has not received adequate attention. At present, investigations into the codon usage preferences of monkeypox virus remain limited and lack systematic, in-depth analysis. Moreover, although codon usage bias has the potential to elucidate key mechanisms underlying virus-host interactions, research focusing on monkeypox virus in this context is still in its preliminary stages.

Therefore, this study aims to investigate the relationship between codon usage bias and host adaptation through an in-depth analysis of the monkeypox virus genome. Utilizing bioinformatics and molecular biology

techniques, we will systematically examine the codon usage patterns in the IMV form of monkeypox virus and explore their correlations with the virus's replication, transmission, and other biological characteristics. Through this research, we anticipate gaining deeper insights into the mechanisms underlying monkeypox virus-host interactions, thereby providing novel scientific evidence and strategies for the prevention and control of monkeypox outbreaks. Moreover, these findings are expected to enhance our understanding of other zoonotic viruses and serve as a valuable reference for addressing potential epidemic threats in the future.

II. Materials and Methods

2.1 Materials

A comprehensive search was conducted in the GenBank database of the National Center for Biotechnology Information (NCBI) using the query term "monkeypox virus complete" yielding 2,039 complete genome sequences of monkeypox virus (MPXV). To ensure the inclusion of sequences derived exclusively from human hosts (Homo sapiens), each entry's host information was meticulously reviewed. Subsequently, custom scripts were employed to extract the coding sequences of 11 IMV genes(Table 1): A9, A13L, A14, A21, A30, E6, I2, I5, J1, L1R, and P21.Codon usage characteristics of these genes were analyzed using CodonW software. To ensure the reliability and accuracy of the results and maintain a high-quality dataset, sequences with anomalous ENC values (a total of 75 fragments from A13L) were excluded from further analysis. The raw data will be provided in the supplementary file.

Table 1: Sequence information

Protein Name	Sequence Tagging
A9	IMV membrane protein A9
A13L	IMV membrane protein A13L
A14	IMV membrane protein A14
A21	IMV membrane protein A21
A30	IMV membrane protein A30
E6	IMV membrane protein E6
I2	IMV membrane protein I2
I5	IMV membrane protein I5
J1	IMV membrane protein J1
L1R	IMV membrane protein L1R
P21	IMV membrane protein P21

This study considered only the 59 synonymous codons encoding 18 amino acids, excluding codons for methionine, tryptophan, and the three stop codons (TAA, TAG, and TGA), as these codons do not exhibit usage bias[12].Consequently, these five codons were not considered in the final analysis.

2.2 Composition Analysis

The compositional characteristics of the coding sequences for the 11 IMV genes were analyzed to include the following: (i). The overall GC content;(Ii). The frequency of each nucleotide at the third position of synonymous codons ($A3_S$%, $C3_S$%, $G3_S$%, and $T3_S$%);(iii). The frequency of G+C at the third position of synonymous codons ($GC3_S$%). The nucleotide composition, ENC, RSCU, and CAI values for each of the 11 genes were calculated using CodonW version1.4.2, developed by J. Peden.

2.3 Analysis of the Effective Number of Codons

ENC is a metric that describes the degree to which codon usage deviates from random selection. Unlike comparing the frequency of specific codons, ENC reflects the uneven usage preferences of synonymous codons within codon families[13]. The ENC value ranges from 20 to 61, where a value of 20 indicates extreme bias, and 61 signifies relatively balanced usage among synonymous codons. Generally, ENC values close to or equal to 35 are considered indicative of moderate codon usage bias[14].

ENC is a dimensionless scalar commonly employed in genomics and bioinformatics as it is a purely calculated metric. The formula for ENC is as follows:

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

$$F = \frac{n\sum_{i=1}^{k} p_i^2 - 1}{n-1}, \quad n > 1, \quad p_i = \frac{n_i}{n} \tag{2}$$

Where, n represents the total number of codons used in the gene, k denotes the number of synonymous codons,and p_i refers to the frequency of the i^{th} codon(n_i/n).The ENC value is influenced by factors such as gene length and amino acid composition, providing critical insights into the codon usage patterns and their potential biological significance.

2.4 Relative Synonymous Codon Usage Analysis

RSCU is employed to assess differences in codon usage patterns between the studied genes and the human genome[15]. The calculation of RSCU values follows the algorithm introduced by Sharp P.M. et al. in 1986.

For synonymous codons, an RSCU value greater than 1 indicates a high-frequency codon, signifying a relatively higher usage frequency of that codon. Conversely, an RSCU value less than 1 suggests lower relative usage[12]. Within coding sequences, codons with RSCU values > 1.6 are considered over-represented, while those with RSCU values <0.6 are deemed under-represented. An RSCU value equal to 1 implies no bias in codon usage[16].

The formula for calculating RSCU is as follows:

$$RSCU_{ij} = \frac{x_{ij}}{\frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}} \tag{3}$$

In the formula, x_{ij} refers to the number of times the j^{th} codon is used to encode the i^{th} amino acid, while n_i represents the total number of synonymous codons available for encoding the i^{th} amino acid.

2.5 Codon Adaptation Index

CAI is a method used to measure how closely the codon usage of a gene aligns with a reference set of genes (typically highly expressed genes) in the host organism. The CAI value ranges from 0 to 1, where a value closer to 1 indicates that the gene's codon usage is highly adapted to the codon preferences of the host[17].The CAI values were calculated using the CodonW v1.4.2 software with the reference organism as Homo sapiens (human). Those sequences with higher CAI values were chosen over the lower CAI values. It also indicates that the frequently used codons will preferably get adapted to their host. The derivation of the CAI calculation formula is as follows:

w_{ij} (The relative adaptiveness of a codon): The relative adaptiveness of a codon.

$$w_{ij} = \frac{RSCU_{ij}}{RSCU_{i\max}} = \frac{x_{ij}}{x_{i\max}} \tag{4}$$

In the above formula, $RSCU_{i\max}$ and $x_{i\max}$ represent the *RSCU* value and the *x* value, respectively, of the codon with the highest usage frequency encoding the i^{th} amino acid.

$$CAI = \left(\prod_{K=1}^{L} w_k\right)^{\frac{1}{L}} \tag{5}$$

Here, L refers to the number of codons used in the gene.

2.6 The Role of Mutational Pressure and Natural Selection in Shaping IMV Codon Preference

ENC plot analysis is an intuitive graphical method used to reflect the extent to which mutational pressure influences codon usage bias[18]. In this analysis, the GC3s content is plotted on the x-axis, and the ENC value is plotted on the y-axis, forming an ENC-GC3s scatter plot. If the corresponding points lie close to the expected curve, mutational pressure is the primary factor influencing codon diversity. Conversely, if the points are significantly below the expected curve, natural selection has a greater impact on codon preference. The expected ENC curve is calculated as follows:

$$ENC^{\exp ected} = 2 + GC_{3s} + \frac{29}{\left[GC_{3s}^2 + \left(1 - GC_{3s}\right)^2\right]} \tag{6}$$

Neutral plot analysis is used to examine the correlation between the GC content at the first and second codon positions (GC12) and the GC content at the third codon position (GC3)[19]. In this analysis, the x-axis represents the GC3 content, while the y-axis represents the GC12 content of each gene in the genome. Using MATLAB software, GC3 and GC12 values are calculated and plotted as a two-dimensional scatter plot, followed by linear fitting. When the regression coefficient is close to 1, it indicates that mutational pressure is the primary factor influencing codon preference. Conversely, when the regression coefficient is close to 0, it suggests that natural selection is the dominant factor affecting codon preference[20].

PR2 plot analysis can be used to determine the magnitude and direction of gene bias[21]. This method examines the influence of mutational pressure and natural selection on codon preference based on the GC bias values of important genes. The x-axis represents the GC bias value of the target gene [G3/(G3 + C3)], while the y-axis represents the AT bias value [A3/(A3 + T3)]. PR2 plot analysis is then performed. The central coordinate (0.5, 0.5) represents the position where A = T and C = G, indicating that the target gene at this position is unaffected by mutation or selection. If points are evenly distributed across the four quadrants, codon usage bias is primarily caused by mutational pressure. Uneven distribution suggests that other factors, such as natural selection, also play a role [22].

III. Results

3.1 Nucleotide Composition Analysis Results

The GC3s content is a valuable indicator of the degree of base composition bias, representing the frequency of G + C nucleotides at the synonymous third codon positions, excluding Met, Trp, and stop codons[23]. Among the base compositions of the 11 genes classified from the IMV membrane protein gene (Figure 1), the highest GC content is observed in gene A14, with a value of 39.39% ± 0.0027 (mean ± SD), while the lowest is found in gene I2, with a value of 27.03% ± 0.0077. The SD values of the 11 genes are close to the mean GC content, demonstrating strong stability. Furthermore, the GC3s content of all genes does not exceed 40%. It is evident that the IMV membrane protein genes of monkeypox virus are rich in A/T bases.An analysis of the third nucleotide composition of synonymous codons (T3s, C3s, A3s, G3s) revealed that codons ending with A were the most prevalent, accounting for 8 types, followed by codons ending with T, which made up the remaining 3 types. Notably, in gene A30, the A3s content reached 58.60% ± 0.0546, while in genes J1 and L1R, the A3s and T3s contents were approximately equal.

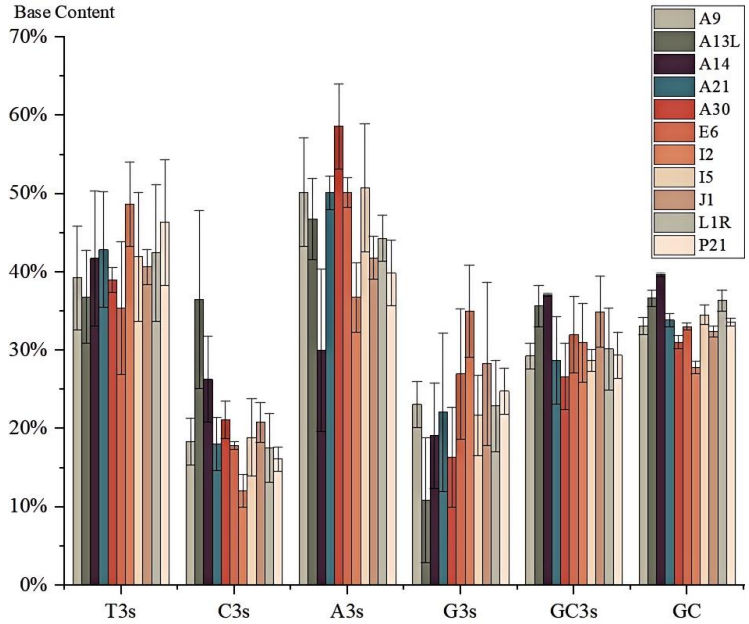

Figure 1: Codon Usage Metrics (T3s, C3s, A3s, G3s, GC3s, GC Content) of 11 Genes in IMV

3.2 Codon Usage Pattern Analysis Results

ENC value is used to quantify codon usage bias in each gene. A higher ENC value indicates lower codon usage bias. The ENC analysis of the 11 genes is shown in Figure 2. Among them, most ENC values for gene I5 are around 35. The ENC values for genes A9, A30, and L1R range between 40 and 45, while those for genes A13L, A14, and P21 are approximately 45. For the remaining genes (A21, E6, I2, J1), most ENC values are greater than 45, indicating lower codon usage bias.

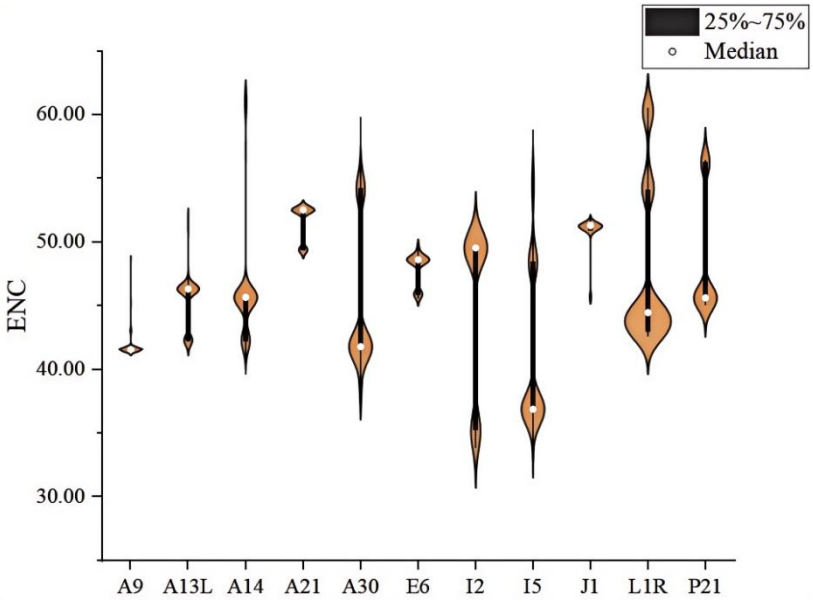

Figure 2: ENC Results of 11 Genes in IMV

3.3 Codon Adaptation Analysis Results

CAI is used to measure the similarity in codon usage between the host and the virus. As shown in Figure 3, the CAI values of these 11 genes are relatively low, with none exceeding 0.35, indicating minimal codon usage bias.Among them, A13L shows relatively higher CAI values compared to other genes, with the median values ranging between 0.30 and 0.325. However, these values are still close to 0, suggesting that they fail to reflect a strong preference for synonymous codons.

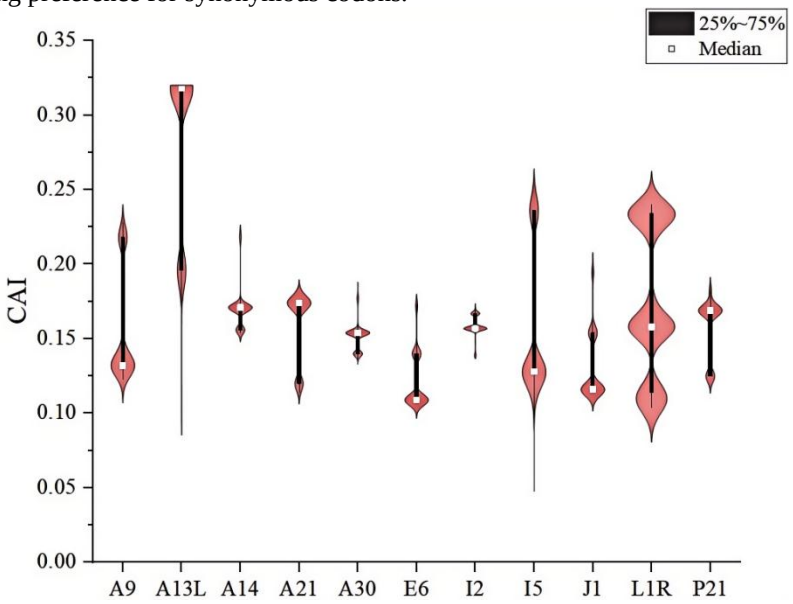

Figure 3: CAI Results of 11 Genes in IMV

3.4 RSCU Analysis Results

This study calculated the differences in RSCU between the genes A9, A13L, A14, A21, A30, E6, I2, I5, J1, L1R, and P21 and those of the host genome to reflect the synonymous codon usage patterns of the genomic coding sequences. A comparative analysis was conducted. In particular, we analyzed the differences in RSCU between the 11 genes of monkeypox virus (IMV) and human genes. Specifically, we computed the differences between the RSCU values of each gene and those of human genes to assess the codon usage preference differences across species.

To facilitate understanding and analysis, we defined the calculation formula for the RSCU differences between each gene and human genes as follows:

△RSCU1=RSCU(A9)-RSCU(human),△RSCU2=RSCU(A13L)-RSCU(human),

△RSCU3=RSCU(A14)-RSCU(human),△RSCU4=RSCU(A21)-RSCU(human),
△RSCU5=RSCU(A30)-RSCU(human),△RSCU6=RSCU(E6)-RSCU(human),
△RSCU7=RSCU(I2)-RSCU(human),△RSCU8=RSCU(I5)-RSCU(human),
△RSCU9=RSCU(J1)-RSCU(human),△RSCU10=RSCU(L1R)-RSCU(human),
△RSCU11=RSCU(P21)-RSCU(human).

The calculation of these differences helps us better understand the codon usage preference of monkeypox virus genes in the process of adaptation to the human host. As shown in Figure 4, the differences in relative synonymous codon usage (ΔRSCU > 1.6\Delta RSCU > 1.6) indicate that the codons significantly overrepresented compared to the host are: CCA, GTA, GGT (A9); AGA, GGA, TCC, GTA, CTA (A13L); GGA, CGT (A14); TTA, GCA (A21); TTA, GGA, GAA, GCA (A30); TTA (E6); CCC, ACT, TTA, GGT, CGT, GTT (I2); ACG, GCA, TTA, CTA, TCA (I5); GGA, TTA (L1R). No significantly overrepresented codons were found for J1 and P21. Most of the optimal codons in these genes tend to end with the base A, and the codon TTA is highly abundant in many genes.

The codons significantly underrepresented compared to the host (ΔRSCU < -1.2) for the 11 genes are as follows: A9 (CTG, GTG, GCC, CCT, CAG), A13L (GTG, GCC, AGC, CAG, CTG, GGC), A14 (CTG, AGG), A21 (AGC, ATC), A30 (CTG, GTG, ACC, AGC, GGC, CCT), E6 (CTG, GCC, GTG), I2 (AGC, CTG, ACC, GGC, CCT), I5 (GTG, AGA, AGC, GCC), J1 (GCC, GGC, ACC), and L1R and P21 (CTG). This indicates that codons ending with C/G are less frequent, with some even showing values of zero.

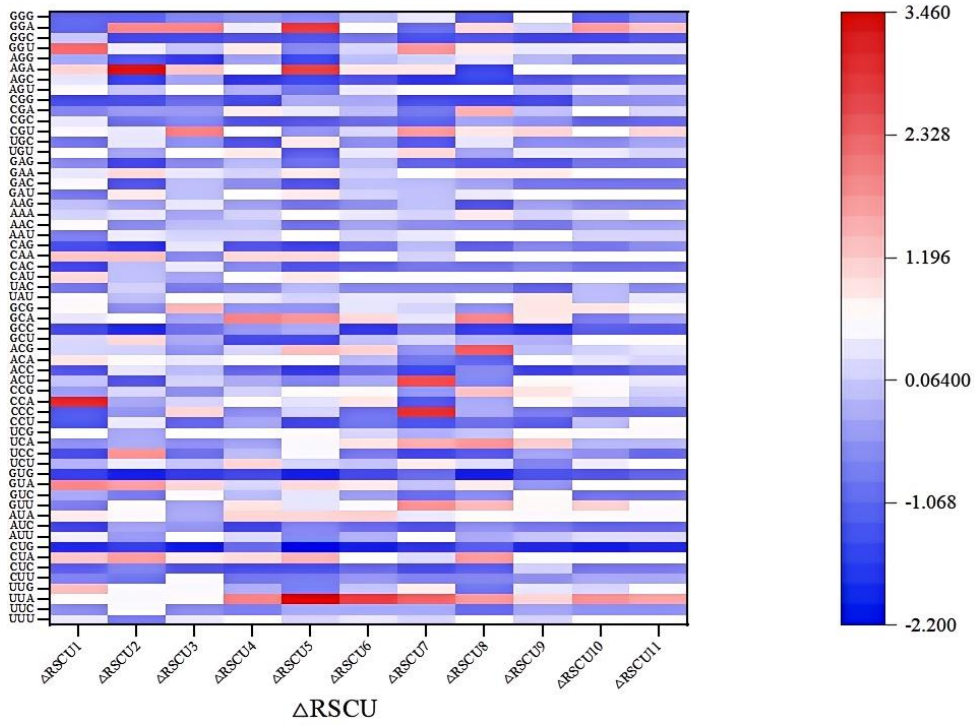

Figure 4: Heatmap of the Differences in RSCU Between 11 IMV Genes and Host Genes

3.5 Influence of Codon Usage Bias

To determine the influence of codon usage bias on the IMV membrane protein genes of monkeypox virus, ENC plots were constructed for the sequences of 11 genes from IMV, as shown in Figure 5. The results indicate that most points for the 11 genes are distributed below the standard curve. All points for the E6 gene are located below the standard curve, while a few points for the I2 and J1 genes lie on the curve. The remaining points are all below the expected ENC curve. Among these genes, partial sequences of A9, A13L, A21, A30, I2, and L1R are located near the curve, while others deviate significantly, indicating that these genes are influenced by both mutational pressure and natural selection. The points for E6, I5, J1, and P21 are closer to the standard curve, suggesting that their codon usage is primarily affected by mutational pressure. In contrast, the points for the A14 gene deviate substantially from the standard curve, indicating that its codon usage is influenced by mutational pressure but predominantly by natural selection.

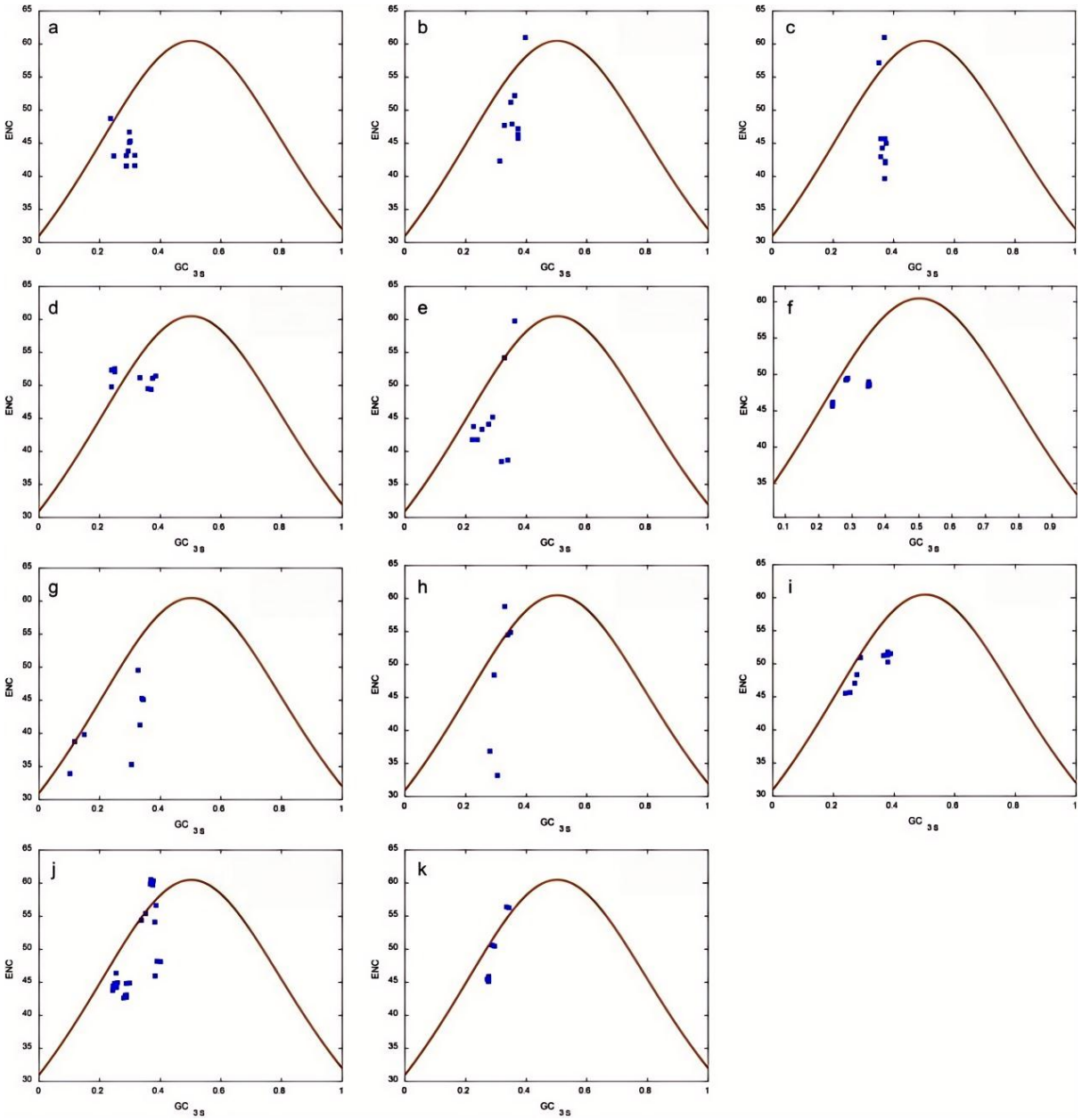

Figure 5: ENC-GC3 Plot Analysis of 11 Genes in IMV (The solid line represents the expected curve when codon usage bias is solely influenced by mutational pressure) (a)A9; (b)A13L; (c)A14; (d)A21; (e)A30; (f)E6; (g)I2; (h)I5; (i)J1; (j)L1R; (k)P21.

PR2 plot analysis was performed to examine codon usage bias in the 11 genes of IMV in monkeypox (Figure 6). The plot is centered at 0.5 and divided into four quadrants. It evaluates the impact of mutation and natural selection on codon usage bias. When the third nucleotide of codons satisfies A = T and C = G, it indicates that no selective effects exist on the complementary DNA strands, and mutation is the sole factor influencing codon bias[24]. The results showed an irregular distribution of these 11 genes across the four quadrants. The I2, J1, L1R, and P21 genes were primarily distributed in the fourth quadrant. The A9, E6, and I5 genes were mainly located in the second quadrant, followed by the fourth quadrant. The A13L and A30 genes were predominantly distributed in the first and fourth quadrants, while the remaining genes were concentrated in the third and fourth quadrants. These findings indicate that the codon usage bias in these 11 genes results from the combined effects of mutation and natural selection.

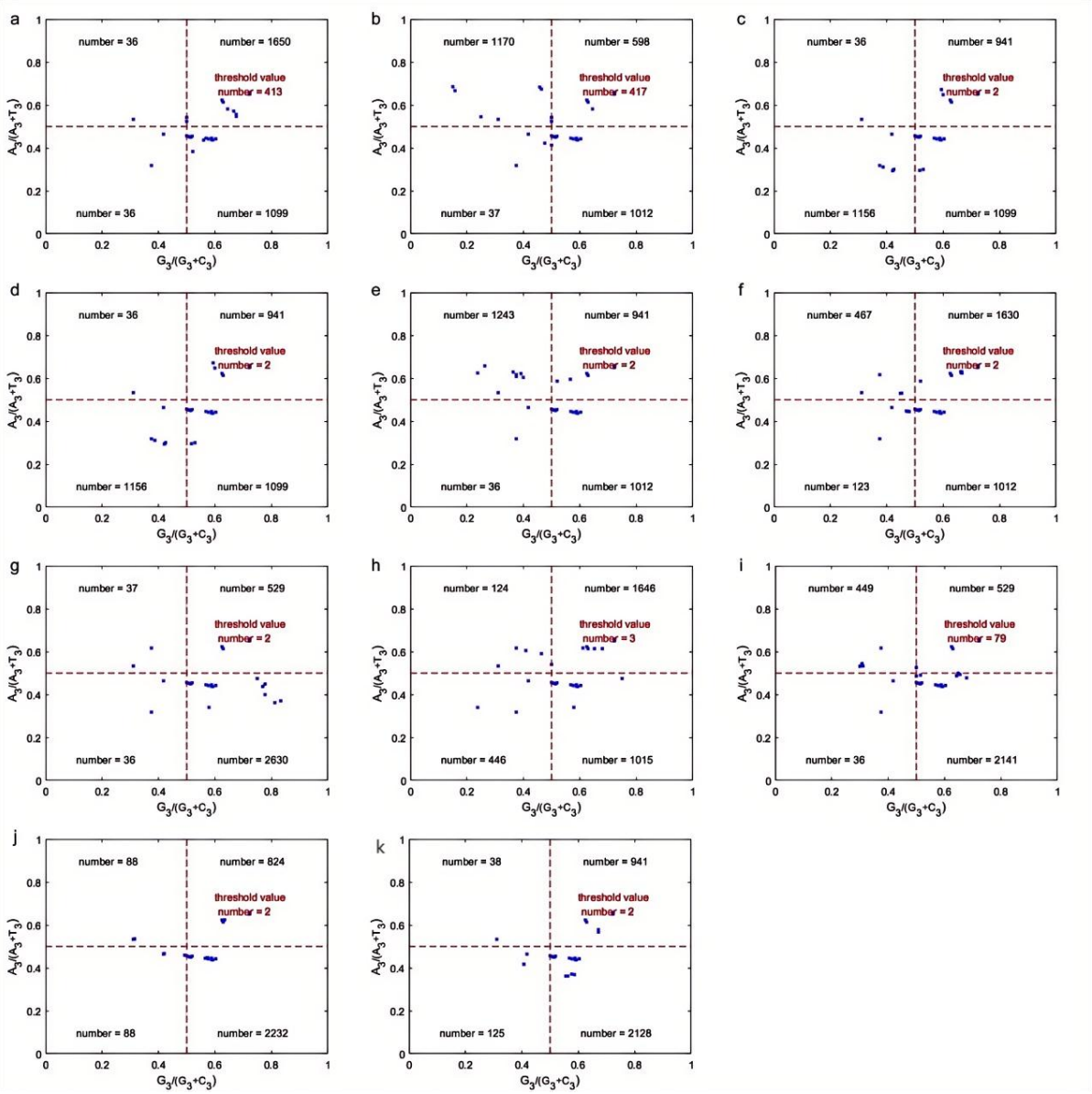

Figure 6: Parity Rule 2 (PR2) Plot Analysis of 11 Genes in IMV
(a)A9;(b)A13L;(c)A14;(d)A21;(e)A30;(f)E6;(g)I2;(h)I5;(i)J1;(j)L1R;(k)P21.

Although ENC-plot can represent the contributions of mutational pressure and natural selection to codon usage preferences, it does not estimate the magnitude of each force. Neutral plot analysis determines the relative contributions of mutational pressure or natural selection to the codon usage patterns of individual genes (Figure 7). A comparison between GC3 and GC12 for all strains revealed a negative correlation between GC3 and GC12 in 11 genes. As the GC3 content in the gene sequence increases, the GC12 content decreases correspondingly. The residual norm values for both range from 1.0796 to 3.5635, indicating relatively large values, which suggests that the model fits the observed data poorly, and the error between the model's predicted values and the actual observations is significant. The regression coefficient for gene A13L is 0.63, indicating that codon usage preference in gene A13L is strongly influenced by mutational pressure. The regression coefficients for genes A14 and A21 are 0.33 and 0.38, respectively, suggesting that the codon usage preferences of these two genes are primarily influenced by natural selection. The regression coefficients for genes L1R and P21 are between 0 and 1, slightly closer to the middle, indicating a dual influence from both mutational pressure and natural selection. These analysis results are consistent with the findings from the ENC-GC3 plot analysis. The residual norm values for the remaining genes are too large, leading to significant discrepancies between the actual and predicted data, and therefore, they are not included in the analysis.

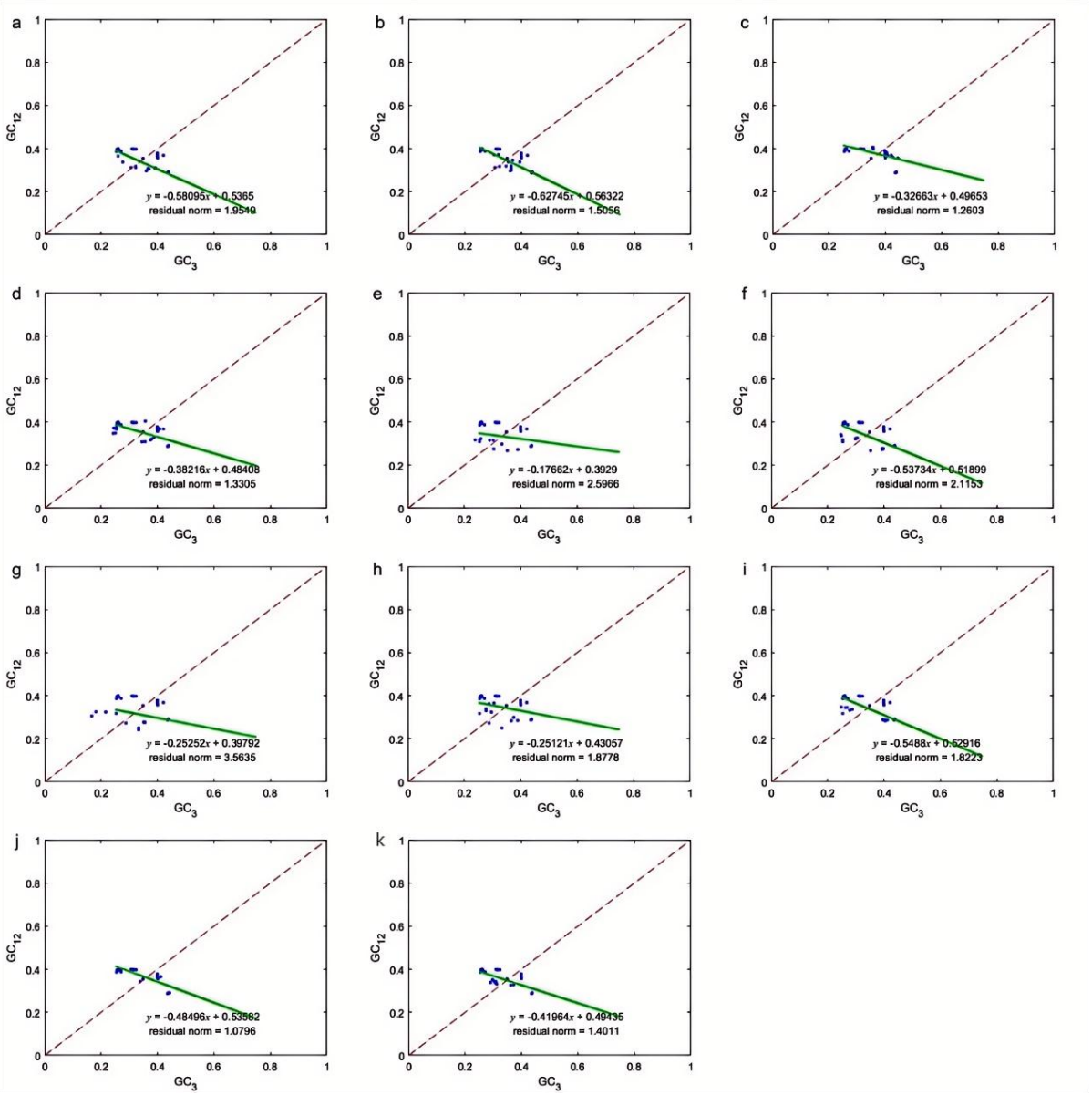

Figure 7: Neutrality Plot Analysis of 11 Genes in IMV
(a)A9;(b)A13L;(c)A14;(d)A21;(e)A30;(f)E6;(g)I2;(h)I5;(i)J1;(j)L1R;(k)P21.

IV. Discussion

Codon preference is formed under the influence of complex factors, resulting from gene mutations and selection[25]. It is also closely related to gene coding structure, function, and gene expression[26], and is influenced by various factors during the evolutionary process. Studying the codon preference of viruses can reveal valuable information about the virus's overall survival, adaptability, and evolution[27], thereby guiding the selection of foreign high-quality expression vectors for target genes or analyzing the gene adaptation mechanisms in certain hosts[28]. Codon bias represents the frequency differences in synonymous codon usage within an organism. In various species in nature, codon bias partially reflects interspecies differences[29]. Given the scarcity of research on the codon preference of monkeypox virus genes and the advantages of codon bias in studying viral gene codon usage patterns and host adaptability, this study conducts an in-depth analysis of the codon preference of 11 gene sequences of the mature virus IMV within monkeypox cells. The study explores the differences in the usage patterns of codons in monkeypox virus genes, the influencing factors, and host adaptation mechanisms, aiming to provide insights into understanding viral infection mechanisms, disease progression, and vaccine development.

This study employed a systematic approach to obtain the complete genome sequence of the monkeypox virus, ensuring that the selected sequences were derived from a human host. Additionally, a script was used to extract the gene sequences of the mature virus IMV within the cells, enhancing both efficiency and accuracy through this automated extraction method. The obtained sequences were classified into 11 distinct gene sequences: A9, A13L, A14, A21, A30, E6, I2, I5, J1, L1R, and P21. Comparative analysis was performed on these 11 genomic sequences using various analytical methods, including the codon adaptation index, relative synonymous codon usage, effective number of codons, and the roles of mutational pressure and natural selection in shaping gene codon preference. These methods provide a comprehensive evaluation of the genetic encoding features and patterns in the monkeypox virus IMV membrane protein genome, offering significant insights for a deeper understanding of its genetic mechanisms.

This study determined the base composition of the coding regions of 11 genes to assess the potential impact of base composition on codon usage bias. Additionally, GC composition plays an important role in the evolutionary process[30]. It was observed that the nucleotide compositions of these genes showed significant similarities, with the GC content of the first three bases being very similar and all below 40%, indicating a preference for codons ending in A/T. The findings are consistent with previous studies, suggesting that the MPXV coding sequences are A/T-rich. Notably, the A/T content of the WA-Outbreak2022 strain was also slightly higher[31].

Subsequently, the ENC values of the 11 genes were analyzed. The ENC value is an important parameter and plays a key role in explaining codon usage patterns[32]. The violin plot analysis of the ENC values (Figure 2) revealed that the ENC values of all gene sequences were generally greater than 35, indicating relatively low codon usage bias. This finding is consistent with previous studies[33]. Such low codon usage bias may facilitate the survival and replication of the virus within the host by minimizing competition for protein synthesis resources between the host and the internal virus[34].

Next, this study determined the CAI values of the 11 genes and plotted a violin plot (Figure 3). CAI is one of the key parameters used in research analysis and measures the degree of bias toward highly expressed genes[35], typically ranging from 0 to 1. In this study, the CAI values of these genes were relatively low, none exceeding 0.35, indicating very low usage bias. This suggests that the expression levels of these genes within the organism are not high, meaning that the codon usage of monkeypox genes is not highly similar to that of highly expressed genes in the host. Consequently, the adaptability and expression potential of these genes within the host are low, which also presents challenges for further investigation of their biological functions in the organism[36].

Additionally, the differences in the RSCU values between the 59 synonymous codons of the 11 genes and the host were calculated. The differences in the RSCU values between monkeypox virus and human genes reflect their evolutionary characteristics. The heatmap analysis of △RSCU values (Figure 4) identified certain codons (CCA, AGA, GGA, TTA, GAA, CCC, ACT, ACG) that were significantly overrepresented in the virus compared to the host, while other codons (CTG, GTG, AGC, GCC, AGG, ATC, ACC) were underrepresented. It was observed that most codons ending in A were overrepresented, while those ending in G/C were underrepresented, indicating that codons with A at the end are generally the preferred codons in monkeypox virus. These results suggest that selective pressure from the host may also influence the codon usage patterns in monkeypox virus.

From the nucleotide content and RSCU analysis, this study suggests that the selection of preferred codons is typically constrained by composition, which determines the presence of mutational pressure. However, mutational pressure is not the only factor associated with gene codon usage patterns, as there are variations in codons among different genes in the IMV membrane protein genome. For instance, the genes I2 and I5 predominantly feature the codons CCC, ACG, and ACT.

Finally, to further explore the factors influencing the formation of codon preference in the IMV membrane protein genes, this study performed comparative analyses on the 11 gene sequences using ENC plots, PR2 plots, and neutral plots. The ENC plot showed a comparison between GC3 values and ENC values, revealing codon usage bias in the IMV membrane protein genome. In the ENC plot, most data points for the 11 genes were located below the standard curve, with only a few points above the expected curve, indicating that the codon usage pattern is shaped by a combination of mutational pressure and natural selection. A PR2 plot was constructed with A3/(A3+T3) as the vertical axis and G3/(G3+C3) as the horizontal axis. The analysis revealed that the usage of AT and GC codons at the third codon position was inconsistent, suggesting that, in addition to mutations, natural selection within the monkeypox virus cells may also influence gene codon usage patterns. The neutral plot explained the relationship between GC12 and GC3 and was used to assess the impact of the mutation-selection balance on the formation of codon preference. In this study, it was found that GC3 and GC12 were negatively correlated in the 11 genes, with most genes being the result of the combined action of natural selection and mutational pressure. Only a few genes, such as A14 and A21, had regression coefficients of 0.33 and 0.38, respectively, indicating that the codon usage bias in these genes is more heavily influenced by natural

selection. The results showed that the outcomes of the three plotting methods were very similar, suggesting that the formation of codon preference in monkeypox virus genes is the result of the combined effects of natural selection and mutational pressure.

V. Conclusion

This study provides a comprehensive analysis of the IMV membrane protein genes in MPXV from the perspective of codon preference, focusing on 11 distinct gene sequences within these genes. The results indicate that the codon usage bias in the monkeypox virus IMV membrane protein genes is relatively weak, with both mutational pressure and natural selection influencing codon preference. Notably, the expression levels of these genes in the species are low, and there is a preference for codons ending in A/T.

Codon preference is just one of many factors influencing gene expression, and codon usage itself is impacted by a variety of factors within biological systems. Therefore, understanding codon usage bias is not as simple as altering DNA sequences or codons, but is highly complex due to the difficulty in identifying or measuring the relative influence of various factors or components. Despite these limitations, the information obtained in this study provides valuable insights for understanding viral infection mechanisms and disease progression.

Acknowledgments

The authors thank the Science and Technology Department of Shaanxi Province for funding - key research and development project from the Science and Technology Department of Shaanxi Province (project code: 2023-YBGY-261).

References

[1] Adler H, Gould S, Hine P, et al. Clinical features and management of human monkeypox: a retrospective observational study in the UK[J]. The Lancet Infectious diseases, 2022, 22(8): 1153-1162.
[2] Priyamvada L, Satheshkumar PS. Variola and monkeypox viruses(Poxviridae)[M]. Reference Module in Biomedical Sciences, 2014.
[3] Mitjà O, Ogoina D, Titanji BK, et al. Monkeypox[J]. The Lancet Infectious diseases, 2023, 401(10370): 60-74.
[4] Letafati A, Sakhavarz T. Monkeypox virus: A review[J]. Microbial Pathogenesis, 2023, 176: 106027.
[5] Alakunle E, Moens U, Nchinda G, et al. Monkeypoxvirus in Nigeria:infection biology,epidemiology,and evolution[J]. Viruses, 2020, 12(11): 1257.
[6] Papukashvili D, Rcheulishvili N, Liu C, et al. Strategy of developing nucleicacid-based universal monkeypox vaccine candidates[J]. Frontiers in Immunology, 2022, 13: 1050309.
[7] Velavan TP, Meyer CG. Monkeypox 2022 outbreak: An update[J], Tropical medicine & international health, 2022, 27(7): 604-605.
[8] Karagoz A, Tombuloglu H, Alsaeed M, et al. Monkeypox (mpox) virus: Classification, origin, transmission, genome organization, antiviral drugs,and molecular diagnosis[J]. Journal of Infection and Public Health, 2023, 16(4):531-541.
[9] Lee W, Kim YJ, Lee SJ, et al. Current Status of Epidemiology, Diagnosis, Therapeutics, and Vaccines for the Re-Emerging Human Monkeypox Virus[J]. Journal of microbiology and biotechnology, 2023, 33(8): 981-991.
[10] S Celina S, Černý J. Genetic background of adaptation of Crimean-Congo haemorrhagic fever virus to the different tick hosts[J]. PLOS ONE, 2024, 19(4): e0302224.
[11] Kaushik R, Kumar N, Yadav P, et al. Comprehensive Genomics Investigation of Neboviruses Reveals Distinct Codon Usage Patterns and Host Specificity[J]. Microorganisms, 2024, 12(4): 696.
[12] Rahman SU, Rehman HU, Rahman IU, et al. Analysis of codon usage bias of lumpy skin disease virus causing livestock infection[J]. Frontiers in Veterinary Science, 2022, 9: 1071097.
[13] Wu, X.M., Wu, S.F., Ren, D.M., et al. Methods for analyzing codon bias and related research progress[J]. Hereditas (Beijing), 2007, 29(4): 420–426.
[14] Wright F. The“effective number of codons”used in a gene[J]. Gene, 1990, 87(1): 23-29.
[15] Xu G, Zou JW, Yang KK, et al. Non-adaptive evolution in codon usage of human-origin monkeypox virus[J]. Comparative Immunology, Microbiology and Infectious Diseases, 2023, 100: 102024.
[16] Patil SS, Indrabalan UB, Suresh KP, et al. Analysis of codon usage bias of classical swine fever virus[J]. Veterinary World, 2021, 14(6): 1450-1458.
[17] Sharp PM, Li WH. The codon adaptation index ameasure of directional synonymous codon usage bias,and its potential applications[J]. Nucleic Acids Research, 1987, 15(3): 1281-1295.
[18] Hussain S, Rasool ST, Pottathil S. The Evolution Of Severe Acute Respiratory Syndrome Coronavirus-2 during Pandemic And Adaptation to the Host[J]. Journal of Molecular Evolution, 2021, 89(6): 341-356.
[19] Sueoka N. Directional mutation pressure and neutral molecular evolution[J]. Proceedings of the National Academy of Sciences of the United States of America, 1988, 85(8): 2653-2657.
[20] Kawabe A, Miyashita NT. Patterns of codonusage bias in three dicot and four monocot plantspecies[J].Genes and Genetic Systems, 2003, 78(5): 343-352.
[21] Wu P, Luo YY, Xiao WQ, et al. Comprehensive analysis of codon bias in 13 Ganoderma mitochondrial genomes[J]. Frontiers in Microbiology, 2023, 14: 1170790.
[22] Li YF, Luo CY, Shi ZF, et al. Codon usage bias and evolutionary analysis of the SARS-CoV-2 Omicron variant[J]. Chinese Journal of Zoonoses, 2022, 38(07): 559-576.
[23] Wang HJ, Liu SQ, Zhang B, et al. Analysis of Synonymous Codon Usage Bias of Zika Virus and Its Adaption to the Hosts[J]. PLOS ONE, 2016, 11(11): e0166260.
[24] Lu Qifeng, Luo Wenhua, Huang Zhihuan. Analysis of codon usage bias in chloroplast genomes of two species of Firmiana[J]. Guihaia, 2020, 40(2): 173-183.

[25] Wong GK, Wang J, Tao L, et al. Compositional gradients in gramineae gene[J]. Genome Research, 2002, 12(6): 851-856.
[26] Chiapello H, Lisacek F, Caboche M, et al. Codon usage and gene function are related in sequences of Arabidopsis thaliana[J]. Gene, 1998, 209(1-2): GC1-GC38.
[27] Yao HP, Chen MY, Tang ZZ. Analysis of Synonymous Codon Usage Bias in Flaviviridae Virus[J]. BioMed Research International, 2019, 2019: 5857285.
[28] Sun SH, Chen ZF, Guo ZX, et al. Study on Codon Bias of N Gene of Three Swine Coronaviruses and Host Adaptation Mechanisms[J]. China Animal Husbandry & Veterinary Medicine, 2023, 50(2): 2403-2413.
[29] Zhang Z, Zhang GS, Zhang XK, et al. Analysis of Codon Usage Bias in Chloroplast Genomes of Hyssopus cuspidatus and Hyssopus officinalis[J]. Molecular Plant Breeding, 2024, 1-14.
[30] Yi SK, Li YH, Wang WM. Selection shapes the patterns of codon usage in three closely related species of genus Misgurnus[J]. Genomics, 2018, 110(2): 134-142.
[31] Zhoqqu JL, Wang XJ, Zhou Z, et al. Insights into the Evolution and Host Adaptation of the Monkeypox Virus from a Codon Usage Perspective: Focus on the Ongoing 2022 Outbreak[J]. International Journal of Molecular Sciences, 2023, 24: 11524.
[32] Li G, Shi L, Zhang L, et al. Componential usage patterns in dengue 4 viruses reveal theirbetter evolutionary adaptation to humans[J]. Frontiers in Microbiology, 2022, 13: 935678.
[33] Yang MQ, Shan J, Li YF, et al. Analysis of Codon Usage Bias Evolution in Monkeypox Virus[J]. Chinese Journal of Parasitic Biology, 2024, 19 (2): 125-129.
[34] Nasrullah I, Butt AM, Tahir S, et al. Genomic analysis of codon usage shows influence of mutation pressure,natural selection,and host features on Marburg virus evolution[J]. BMC Evolutionary Biology, 2015, 15: 174.
[35] De Mandal S, Mazumder TH, Panda AK, et al. Analysis of synonymous codon usage patterns of HPRT1 gene across twelve mammalian species[J]. Genomics, 2020, 112(1): 304-311.
[36] Dai KY, Yang ZF, Guo YX. Analysis of Codon Usage Bias of Toll - like Receptor 5 Gene in 15 Species[J]. China Animal Husbandry & Veterinary Medicine, 2024, 51(1): 11-22.