Recep Tayyip Erdoğan Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 5(1) 32-48 2024
Recep Tayyip Erdogan University Journal of Science and Engineering, 5(1) 32-48 2024
ISSN: 2687-2315 / E-ISSN: 2757-7686

A Comprehensive Comparative Analysis on the Codon Usage Bias of DNA Polymerase Genes in Invertebrate Iridescent Viruses

Yeşim Aktürk Dizman*

*Recep Tayyip Erdogan University, Faculty of Arts and Sciences, Department of Biology, 53100, Rize, Türkiye

*Sorumlu Yazar/Corresponding Author
E-mail: yesim.akturk@erdogan.edu.tr

Araştırma Makalesi/Research Article
Geliş Tarihi/Received:08.01.2024
Kabul Tarihi/Accepted: 27.02.2024

ABSTRACT

Invertebrate iridescent viruses (IIVs) are classified as double-stranded DNA viruses within the Iridoviridae family. IIVs are viruses that infect invertebrate hosts, causing symptoms that vary in intensity from slight reductions in host fitness to systemic disease. Numerous earlier investigations have provided insights into the genomic, proteomic, and transcriptional analyses of invertebrate iridescent viruses. However, the codon usage bias (CUB) of IIVs has yet to be fully understood. In order to gain a more profound insight into the evolutionary features of IIVs, an extensive analysis of the codon usage patterns in the DNA polymerase genes (DNA pol genes) of 12 invertebrate iridescent viruses were conducted. The analysis of both nucleotide composition and relative synonymous codon usage (RSCU) indicated a higher prevalence of AT-ended codons in the DNA pol genes of IIVs. Additionally, a low codon usage bias was determined from the effective number of codons (ENC) value. Analyses of the ENC-GC3s plot, neutrality plot, and parity rule 2 plot illustrated that the codon usage patterns in IIVs DNA pol genes were influenced by both natural selection and mutational pressure. This investigation holds significance as it has delineated the codon usage patterns within the DNA pol genes of IIVs and has furnished crucial data for a foundational study of their evolutionary aspects.

Keywords: DNA polymerase, Invertebrate iridescent viruses, Codon usage bias, Mutation pressure, Natural selection.

Invertebrate Iridescent Virüslerdeki DNA Polimeraz Genlerinin Kodon Kullanım Eğilimi Üzerine Ayrıntılı Karşılaştırmalı Bir Analiz

ÖZ

Invertebrate iridescent virüsler (IIV'ler), Iridoviridae familyası içerinde yer alan çift sarmallı DNA virüsleri olarak sınıflandırılır. IIV'ler omurgasız konakçıları enfekte eden, konak hareketinde hafif azalmalardan sistemik hastalığa kadar değişen yoğunlukta semptomlara neden olan virüslerdir. Daha önce yapılan çok sayıda araştırma, invertebrate iridescent virüslerin genomik, proteomik ve transkripsiyonel analizlerine ilişkin bilgiler sağlamıştır. Bununla birlikte, IIV'lerin kodon kullanım eğilimi henüz tam olarak anlaşılamamıştır. IIV'lerin evrimsel özellikleri hakkında daha derin bir bilgi elde etmek için, 12 invertebrate iridescent virüsün DNA polimeraz genlerindeki kodon kullanım modellerinin kapsamlı bir analizini gerçekleştirildi. Hem nükleotid kompozisyonu hem de göreli sinonim kodon kullanımı (RSCU) analizi, IIV'lerin DNA pol genlerinde AT-uçlu kodonların daha yüksek bir yaygınlıkta bulunduğunu göstermiştir. Ayrıca, etkin kodon sayısı (ENC) değerinden düşük bir kodon kullanım yanlılığı tespit edilmiştir. ENC-GC3s grafiği, nötraliti grafiği ve parite kuralı 2 grafiği analizleri, IIV'lerin DNA pol genlerindeki kodon kullanım modellerinin hem doğal seçilim hem de mutasyon baskısından etkilendiğini göstermiştir. Bu araştırma, IIV'lerin DNA pol genlerindeki kodon kullanım modellerini tanımlaması ve evrimsel yönlerine ilişkin temel bir çalışma için veriler sağlaması açısından önem taşımaktadır.

Anahtar Kelimeler: DNA polimeraz, Invertebrate iridescent virüsler, Kodon kullanım eğilimi, Mutasyon baskısı, Doğal seçilim.

Cite as;
Aktürk Dizman, Y. (2024). A Comprehensive Comparative Analysis on the Codon Usage Bias of DNA Polymerase Genes in Invertebrate Iridescent Viruses, *Recep Tayyip Erdogan University Journal of Science and Engineering*, 5(1), 32-48. Doi: 10.53501/rteufemud.1416072

1. Introduction

Iridoviridae family members have been discovered in ectothermic organisms, encompassing a variety of hosts such as insects, crustaceans, and certain cold-blooded vertebrates. Iridoviruses are categorized into seven genera, which comprise *Iridovirus*, *Chloriridovirus*, *Decapodiridovirus*, *Daphniairidovirus*, *Lymphocystivirus*, *Megalocytivirus*, and *Ranavirus* (Canuti et al., 2022). The insect iridoviruses belonging to the genus *Iridovirus* are noteworthy due to their potential as candidates for biological control. This is because these viruses induce fatal infections in crucial insect species that are considered major pests (Williams, 2008). IIV virions are characterized by an icosahedral capsid enclosing a linear dsDNA molecule (Eaton et al., 2007). The genomic structure of IIVs demonstrates circular permutation and terminal redundancy, a distinctive trait among iridoviruses (Eaton et al., 2010; Schnitzler et al., 1987).

IIVs possess genes that code for DNA polymerase proteins in their genomes (Jakob and Darai, 2002). DNA polymerases play a crucial role in the replication of viral genomes, contributing to the infectious life cycle of diverse viruses. These specialized enzymes are fundamental components of the molecular machinery that facilitates the synthesis of DNA, a process essential for the propagation of genetic information in viral organisms (Chen et al., 2019; Coen et al., 2021). The study of viral DNA polymerases is not only integral to understanding the basic mechanisms underlying viral replication but also holds significant implications for the development of antiviral therapies. In recent years, antiviral drugs have been developed by targeting viral polymerases (Garro and Pungitore, 2019; Bassetto et al., 2019; Sofia et al., 2012). Antiviral medications modify codon usage patterns, leading to diminished HIV-1 fitness (Palanisamy et al., 2017; Hu and Kuritzkes, 2014; Mesplède et al., 2013). A systems biology approach has been utilized to investigate the codon usage profiles of diverse organisms and their adjustments to their respective hosts. Systems biology is additionally valuable for forecasting suitable drugs for a particular disease. Recently, systems biology has been employed for network-based drug repurposing targeting SARS-CoV-2 (Nambou and Anakpa, 2020; Zhou et al., 2020).

Synonymous codons are different codons that specify the same amino acid during protein synthesis. In the genetic code, most amino acids are represented by multiple codons. Since there are 64 possible triplet codons and only 20 standard amino acids, it is possible for multiple codons to code for the same amino acid (Komar, 2016). Codon usage bias (CUB) pertains to the unequal frequency at which various codons are utilized to represent the same amino acid in the genetic code (Behura and Severson, 2013; Iriarte et al., 2021). CUB is shaped by various factors, including mutation, natural selection, tRNA availability, GC content, gene expression levels, and genomic features (Chen et al., 2013; Yannai et al., 2018; Zhong et al., 2007). There are molecular characterization and genetic structure studies of IIVs based on genome analysis (Delhon et al., 2006; Dizman et al., 2016; Yesilyurt et al., 2019). Analysis of codon usage bias in IIVs and Ranaviruses has been reported, employing various protein-coding sequences (Aktürk Dizman, 2023; Tian et al., 2020). Nevertheless, as of now, there have been no investigations into the patterns of codon usage associated with the IIVs DNA pol genes. In this study, a thorough examination of the CUB in the DNA pol genes was conducted. This research offers valuable insights into the determinants affecting the patterns of codon usage detected in the IIVs DNA pol genes which could pave the way to develop novel antiviral therapies against invertebrate vectors transmitting disease to humans in the future.

2. Materials and Methods

2.1. Sequence Data Collection

The DNA pol gene sequences of twelve IIVs were acquired from the NCBI (http://www.ncbi.nlm.nih.gov) (Table 1). These sequences were employed for computing codon usage indices in the analyses. Invertebrate iridoviruses with complete genome sequences and containing the DNA pol gene have been included in the study.

2.2. Compositional Properties Analyses

The CodonW program (https://codonw.sourceforge.net//) was utilized for conducting a nucleotide compositional analysis of the IIVs DNA pol genes. The complete composition of nucleotides (A%, C%, T%, G%, AT%, and GC%), along with the nucleotide content at the 3rd position of synonymous codons (A3s%, C3s%, T3s%, G3s%, and GC3s%), and the GC content at the 1st and 2nd codon positions were computed.

2.3. Relative Synonymous Codon Usage (RSCU) Analysis

RSCU is a measure that indicates the ratio between the observed frequency of a codon and the expected frequency, assuming equal usage of all codons that encode a particular amino acid (Sharp and Li, 1986). RSCU value of 1 signifies equal utilization of the codon, a value below 1 suggests a less favored codon, and a value above 1 indicates a more favored codon. Codons having RSCU values greater than 1.6 are classified as over-represented, whereas those with values below 0.6 are considered under-represented. The CodonW program was utilized for the computation of RSCU values in the DNA pol genes.

2.4. The Effective Number of Codon (ENC) Analysis

ENC serves as a tool for assessing the extent of bias in the codon usage within a given gene. The ENC was determined through the utilization of the CodonW program. The ENC values span from 20 (pointing to a significant bias, only a single synonymous codon for each amino acid is employed) to 61 (indicating no bias, with all synonymous codons being equally utilized) (Smith, 2022). Generally, ENC values below 35 signify a pronounced codon usage bias (Tyagi and Nagar, 2022).

2.5. Codon Adaptation Index (CAI) Analysis

The CAI analysis is applicable for estimating the bias in synonymous codon usage within the nucleic acid sequence that encodes the protein of a given gene. It involves comparing the synonymous codon utilization of a given gene with the synonymous codon occurrence in a reference dataset (Sharp and Li, 1987). The reference dataset for hosts was acquired from the codon usage database (http://www.kazusa.or.jp/codon/). CAI values vary within the range of 0 to 1. An elevated CAI value suggests greater potential for gene expression. CAIcal servers were used for calculating the CAI of DNA pol genes of IIVs (https://ppuigbo.me/programs/CAIcal/) (Puigbò et al., 2008).

2.6. Correspondence Analysis

Correspondence analysis (COA) is a statistical method that examines the associations among several variables. COA was utilized to examine the prevailing patterns of codon usage variation in the DNA pol genes (Suzuki et al., 2008). Every coding sequence for a gene was depicted as a 59-dimensional vector, where each dimension represented the RSCU value of a specific codon (with the exception of AUG, UGG, and stop codons). COA utilizing RSCU values was carried out using CodonW program.

2.7. Aromaticity and Hydropathicity Analysis

The Aroma and Gravy values, which reflect the aromaticity and hydropathicity characteristics of a gene product, can also function as indicators to evaluate the effect of translation or natural selection (Rao et al., 2014). In this research, the Aroma and Gravy values for DNA pol genes in IIVs were detected utilizing the CodonW program.

ble 1. DNA pol genes, virus isolates and host species used in the study

Accession number	Gene	Virus name	Host	Length (bp)
NP_149500.1	IIV6_037L	Invertebrate iridescent virus 6 (IIV6)	*Chilo suppressalis*	3822
YP_009046615	IIV31_001R	Armadillidium vulgare iridescent virus (IIV31)	*Armadillidium vulgare*	4458
YP_654692.1	IIV3_120R	Invertebrate iridescent virus 3 (IIV3)	*Ochlerotatus taeniorhynchus*	3426
YP_004732900.1	IIV9_116R	Wiseana iridescent virus (IIV9)	*Oxycanus dirempta*	3420
YP_009010534.1	IIV25_001R	Invertebrate iridescent 25 (IIV25)	*Simulium vittatum*	3702
YP_009010762.1	IIV22A_001R	Invertebrate iridescent virus 22A (IIV22A)	*Simulium vittatum*	3696
YP_008357299.1	IIV22_001R	Invertebrate iridescent virus 22 (IIV22)	*Simulium vittatum*	4773
YP_009010295.1	IIV30_001R	Invertebrate iridescent virus 30 (IIV30)	*Helicoverpa zea*	4776
YP_009021128.1	AMIV_049	Anopheles minimus iridovirus (AMIV)	*Anopheles minimus*	3363
YP_009552323.1	CQIV_042L	Cherax quadricarinatus iridovirus (CQIV)	*Cherax quadricarinatus*	3150
YP_010084862.1	SHIV_110R	Shrimp hemocyte iridescent virus (SHIV)	*Litopenaeus vannamei*	3150
UUT40408.1	CSAIV_013R	Carnivorous sponge-associated iridovirus (CSAIV)	*Chondrocladia grandis*	2217

2.8. Role of Mutational Pressure and Natural Selection on CUB

2.8.1. ENC-GC3s Plot Analysis

To assess the factors contributing to CUB, a plot of ENC values against GC3s values was created to generate an ENC-GC3s plot, aiming to determine the expected slope. The formula provided below (1) was used to determine the expected ENC values for each GC3s:

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

DNA polymerase genes, in which codon selection is solely determined by mutational bias, will be positioned on or slightly below the anticipated ENC curve. In cases where the codon bias of a gene is impacted by natural selection, the point representing that gene will be distributed well below the curve.

2.8.2. Neutrality Plot Analysis

The neutrality plot is employed for a quantitative assessment of how mutational pressure and natural selection impact the patterns of codon usage within a coding sequence (Zhang et al., 2022). In neutrality plot analysis, a regression line is plotted between GC3s and GC12s. The inclination of this regression line signifies the influence of mutational pressure (Chakraborty et al., 2020). A regression slope of 1 indicates that mutational pressure predominantly influences CUB, while regression curves approaching 0 suggest that natural selection substantially impacts CUB.

2.8.3. Parity Rule 2 (PR2) Plot Analysis

Analysis of parity rule 2 (PR2) was conducted to explore the influences of mutation and natural selection on the codon usage of DNA pol genes in IIVs. A PR2 plot illustrates the GC bias value [G3s/(G3s+C3s)] on the horizontal axis and the AT bias value [A3s/(A3s+T3s)] on the vertical axis. The plot's center, where A=T and G=C, signifies an equal contribution from both mutation and natural selection (Sueoka, 1999).

2.9. Correlation Analysis

The analysis of correlation was employed to assess the connections among variables. The Spearman's rank correlation method was utilized to examine the association between the codon usage patterns and nucleotide content of DNA pol genes in IIVs. All statistical analyses were conducted utilizing OriginPro 9.0.

3. Results

3.1. Nucleotide Composition of the IIVs DNA Pol Genes

The analysis of nucleotide composition in the DNA pol genes of 12 IIVs was conducted to investigate whether compositional constraints might impact the codon usage patterns of these genes (Table 2). The average percentages for A, T, C, and G were 36.36, 28.77, 15.61, and 19.27, respectively (Figure 1A). The nucleotide compositions in the 3rd position of synonymous codons for DNA pol genes in IIVs were also computed. The mean values for A3s%, T3s%, C3s%, G3s%, and GC3s% in these genes were 47.73, 45.19, 20.28, 20.56, and 29.63, respectively (Figure 1B).

Furthermore, ENC values were determined for the DNA pol genes of IIVs (Table 2) to gauge the extent of codon usage bias in these genes. The ENC values for all the examined DNA pol genes varied from 34.60 to 55.23. The ENC value for the DNA pol gene in IIV6 was lower than that in other IIVs, indicating a higher degree of codon usage bias in the DNA pol gene of IIV6 compared to the others. Overall, the mean ENC value was 43.34, suggesting a low level of codon usage bias in DNA pol genes.

Table 2. Nucleotide content of DNA pol genes in IIVs

IIVs DNA pol genes	**A%**	**C%**	**T%**	**G%**	**A3s**	**C3s**	**T3s**	**G3s**	**AT%**	**AT3%**	**GC**	**GC12**	**GC3s**	**ENC**	**CAI**
IIV6_037L	41.30	10.60	30.90	17.20	63.19	6.50	54.53	10.06	72.20	85.40	27.80	34.35	11.60	34.60	0.835
IIV31_001R	36.70	16.00	27.70	19.60	51.68	21.46	42.57	16.60	64.40	69.80	35.60	38.35	28.00	47.02	0.712
IIV3_120R	25.10	24.10	22.30	28.60	14.63	43.38	21.12	53.88	47.40	26.80	52.60	42.40	72.10	44.52	0.538
IIV9_116R	38.50	13.70	31.80	16.00	55.35	12.30	54.06	10.74	70.30	80.40	29.70	34.70	16.70	40.68	0.831
IIV25_001R	38.70	13.00	32.40	15.90	56.54	10.87	55.76	9.34	71.10	82.70	28.90	34.70	14.60	39.23	0.511
IIV22A_001R	40.30	11.20	33.00	15.40	60.71	7.10	57.81	8.71	73.40	86.00	26.60	32.90	11.20	35.40	0.509
IIV22_001R	40.00	11.60	33.00	15.40	59.93	6.62	59.05	7.85	73.00	87.00	27.00	34.05	10.20	35.29	0.506
IIV30_001R	39.80	12.10	32.70	15.40	59.56	7.83	57.54	8.60	72.50	85.70	27.50	34.15	11.70	35.79	0.737
AMIV_049	33.30	18.30	27.40	21.00	38.86	27.51	37.53	30.64	60.70	56.20	39.30	37.00	42.00	55.23	0.479
CQIV_042L	36.70	16.50	26.20	20.60	45.27	26.86	39.24	23.40	62.90	61.10	37.10	36.10	36.20	49.88	0.794
SHIV_110R	36.70	16.50	26.20	20.60	45.27	26.86	39.24	23.40	62.90	61.10	37.10	36.10	36.20	49.88	0.588
CSAIV_013R	29.20	23.70	21.60	25.50	21.80	46.03	23.83	43.50	50.80	33.30	49.20	40.45	65.10	52.57	0.547
Mean	36.36	15.61	28.77	19.27	47.73	20.28	45.19	20.56	65.13	67.96	34.87	36.27	29.63	43.34	0.63
SD	4.888	4.572	4.130	4.304	15.69	14.12	13.33	15.22	8.77	20.91	8.774	2.850	21.449	7.460	0.139

This indicates a predilection for A and T nucleotides at the 3rd codon position. Moreover, the mean AT3 compositions were 67.96%, with the AT composition (65.13%) being significantly higher than the GC composition (34.87%) within the IIVs DNA pol genes (Figure 1C, D). These findings indicate an AT-rich composition, with adenine being the most frequent mononucleotide in DNA pol genes of IIVs.

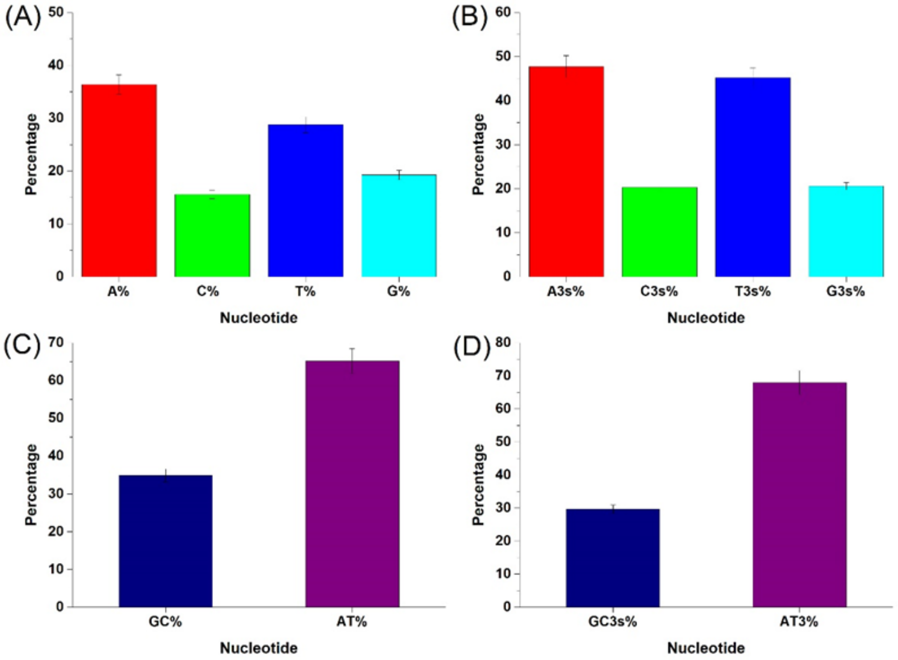

Figure 1. Examination of nucleotide content within DNA polymerase genes of IIVs. (A) The mean values representing the nucleotide compositions of A, T, G, and C. (B) The average frequencies of the third base within codons (A3s, C3s, T3s, and G3s). (C) The mean incidence of GC and AT composition. (D) The mean frequency of GC and AT at the 3rd position within codons.

3.2. RSCU Analysis

The RSCU value, regardless of amino acid content, has been extensively employed to assess the CUB across genes. The mean RSCU values for the DNA pol genes were determined as a result of *in silico* analysis (Table 3). Consistent with the nucleotide composition analysis, 26 out of the 59 synonymous codons were identified as optimal codons (RSCU > 1.0), with 25 of them having A/T endings. Within the category of preferred codons, six codons (TTA, GTT, CCA, ACA, AGA, and GGA) were found to be overrepresented, exhibiting RSCU values exceeding 1.6. It is noteworthy that nearly all of the overrepresented codons were terminated with A. Likewise, 14 codons (TTC, CTA, CTG, GTC, TCC, AGC, CCG, CAG, AAG, GAG, CGC, CGG, GGC, and GGG) with an RSCU value lower than 0.6 were designated as underrepresented, and out of these, 13 were terminated with G or C. The findings suggest the presence of CUB within the DNA pol genes, with a predilection for codons concluding with A or T.

3.3. Codon Adaptation Index Analysis

CAI values were utilized to evaluate the codon usage and adaptive characteristics of DNA pol genes in IIVs to their hosts. In general, there is a consensus that sequences exhibiting higher CAI values indicate a more pronounced level of adaptation to hosts when contrasted with those having lower values. In this context, the CAI values for the DNA pol genes of IIV6 and IIV9 were 0.835 and 0.831, respectively (Table 2), whereas the average CAI values for all DNA pol genes in IIVs were 0.63 (Table 2). These findings suggest that the DNA pol genes of IIV6 and IIV9 exhibit the highest degree of adaptation to their respective hosts.

Table 3. RSCU analysis of IIVs DNA pol genes

AA	Codon	RSCU	AA	Codon	RSCU
Phe (F)	TTT	***1.56***	Ala (A)	GCT	***1.42***
	TTC	0.44		GCC	0.76
Leu (L)	TTA	***2.16***		GCA	***1.17***
	TTG	***1.26***		GCG	0.66
	CTT	1.05	Tyr (Y)	TAT	***1.30***
	CTC	0.62		TAC	0.70
	CTA	0.41	His (H)	CAT	***1.33***
	CTG	0.50		CAC	0.67
Ile (I)	ATT	***1.58***	Gln (Q)	CAA	***1.47***
	ATC	0.63		CAG	0.53
	ATA	0.79	Asn (N)	AAT	***1.28***
Val (V)	GTT	***1.71***		AAC	0.72
	GTC	0.46	Lys (K)	AAA	***1.44***
	GTA	***1.03***		AAG	0.56
	GTG	0.80	Asp (D)	GAT	***1.40***
Ser (S)	TCT	***1.60***		GAC	0.61
	TCC	0.56	Glu (E)	GAA	***1.56***
	TCA	***1.47***		GAG	0.44
	TCG	0.72	Cys (C)	TGT	***1.38***
	AGT	***1.15***		TGC	0.62
	AGC	0.49	Arg (R)	CGT	0.77
Pro (P)	CCT	***1.04***		CGC	0.54
	CCC	0.71		CGA	0.85
	CCA	***1.78***		CGG	0.33
	CCG	0.47		AGA	***2.63***
				AGG	0.88
Thr (T)	ACT	***1.01***	Gly (G)	GGT	***1.28***
	ACC	0.74		GGC	0.53
	ACA	***1.61***		GGA	***1.73***
	ACG	0.64		GGG	0.46

The codons that are predominantly used and their corresponding RSCU values for DNA polymerase genes of the IIVs are emphasized in bold and italicized format. AA: Amino acids.

3.4. Correspondence Analysis

The correspondence analysis was carried out utilizing RSCU values for each of the 59 sense and synonymous codons in every gene. The generated plot illustrates the variation in the distribution of genes and their corresponding codons (Figure 2). It was noted that a majority of the codons were in close proximity to the axes. Consequently, these findings suggest that the pattern of codon usage variations might be influenced by both base distributional constraints and mutational bias.

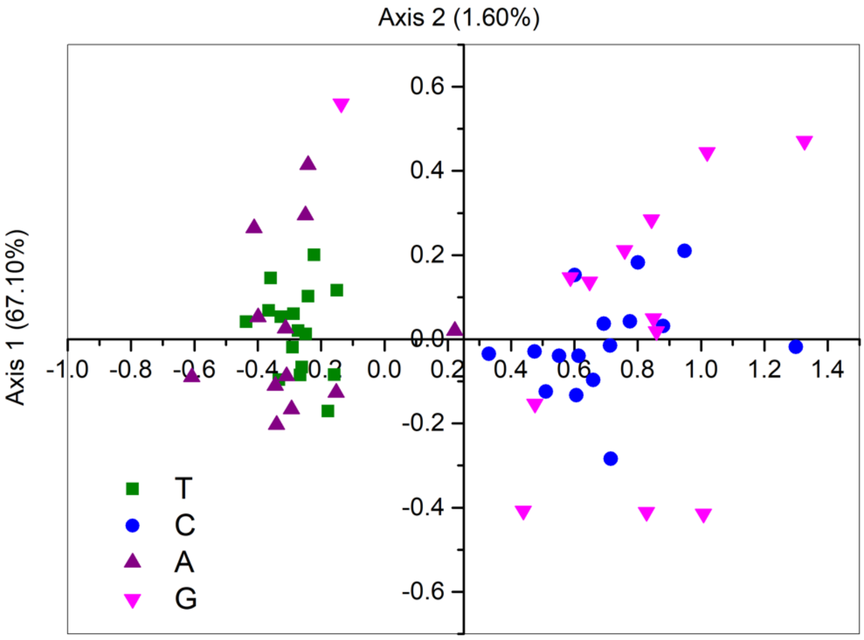

Figure 2. Correspondence analysis of IIVs DNA pol genes. The analysis was carried out utilizing the RSCU values for the 59 synonymous codons. Every symbol in the graph denotes distinct codons concluding with diverse bases.

3.5. Role of Mutational Pressure and Natural Selection on CUB

The construction of the ENC-GC3s plot aimed to investigate the factors that impact CUB in DNA pol genes. The ENC-GC3s plot for DNA pol genes indicated that restricted set of genes displayed ENC values closely grouped around the anticipated ENC curve, while most of the data points were notably situated below the expected curve (Figure 3). This observation suggested that CUB in DNA pol genes of IIVs was not solely influenced by mutational pressure; rather, translational selection and other factors likely contributed to the determination of CUB in these genes.

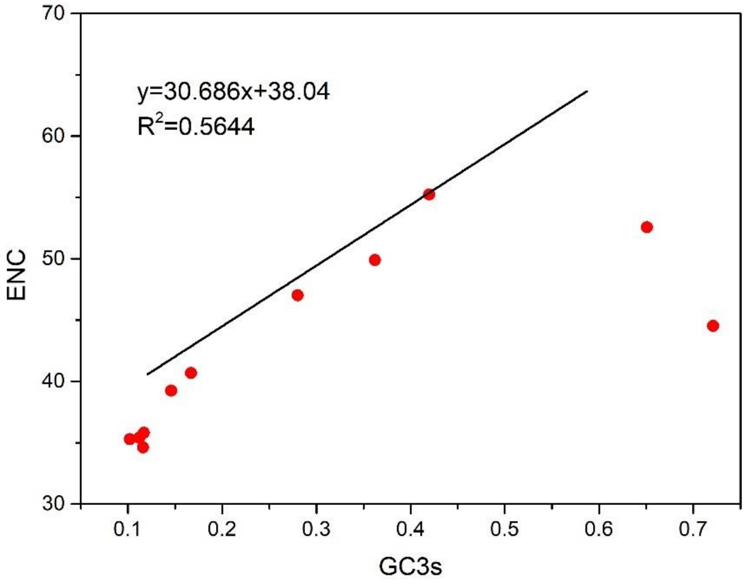

Figure 3. ENC-GC3s plot analysis for the DNA pol genes. The plot showcases the relationship linking the ENC and the GC content found within the 3rd position of synonymous codons. The solid red line indicates the expected ENC values.

The neutrality plot can be used to assess the extent to which natural selection and mutational pressure influence CUB of a gene. In this study, the neutrality plot was created, with GC12s plotted on the Y-axis and GC3s on the X-axis (Figure 4). Furthermore, the DNA pol genes of IIVs displayed regression coefficient value of GC12s on GC3s as 0.1252, signifying a mutation pressure of 12.52%. Consequently, the proportion of natural selection was 87.48%. These findings implied that, in DNA pol genes of IIVs, the impact of natural selection was more prominent than that of mutation pressure.

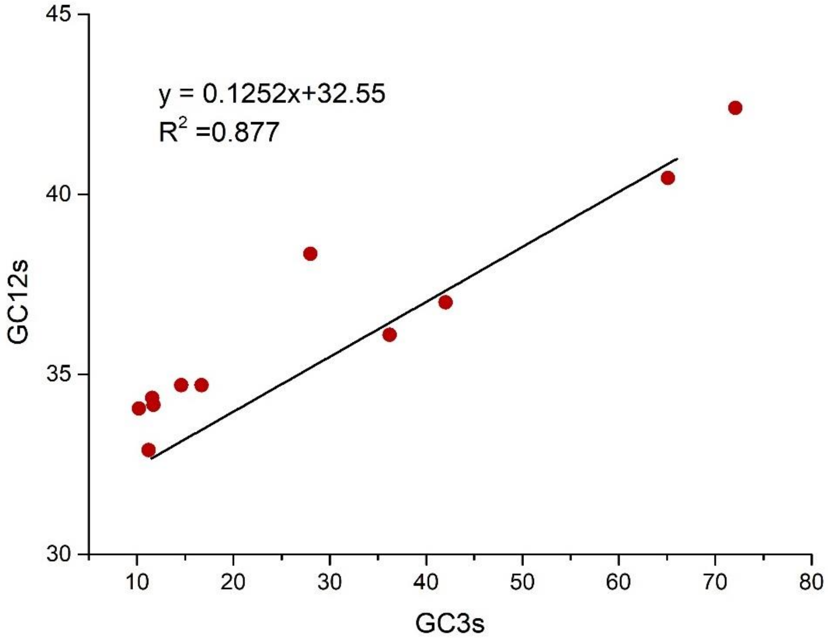

Figure 4. Neutrality plot analysis of IIVs DNA pol genes. The solid line illustrates the linear regression association between GC12s and GC3s.

The PR2-bias plot analysis is frequently utilized to discern the predominant influence of mutation or selection bias on shaping genetic composition. Both axes were aligned to 0.5, dividing the plot into four quadrants. The majority of points were observed to be dispersed in the first and second quadrants of the DNA pol genes in IIVs, suggesting a pronounced inclination towards A over T and G over C in these genes (Figure 5). The results suggested that CUB in DNA pol genes of IIVs was influenced by a blend of mutation along with other elements, such as translational selection.

3.6. Correlation Analysis

Mutation pressure and natural selection are two influential evolutionary forces that have the potential to mold the pattern of codon usage (Chen et al., 2017). A correlation analysis was carried out, examining the relationship between nucleotide compositions at the 3rd position of codons, ENC values, and overall nucleotide compositions, to evaluate the impact of mutation pressure on CUB. Highly significant correlations were noted among these variables (Table 4). These findings suggested that, aside from mutational pressure, other factors also contributed to CUB in the DNA pol genes of IIVs.

To examine the effect of natural selection on CUB in DNA pol genes, a correlation analysis was conducted between Aroma and Gravy values and ENC, GC3s, GC, as well as the values of the first two principal axes (Table 4). Aroma values exhibited no correlation with GC3s, G3s, C3s, A3s, T3s, and ENC values. However, Gravy values displayed a notable positive correlation with Aroma and Axis 2 values. These results suggested that, beyond mutational pressure, natural selection had exerted an influence on CUB of DNA pol genes.

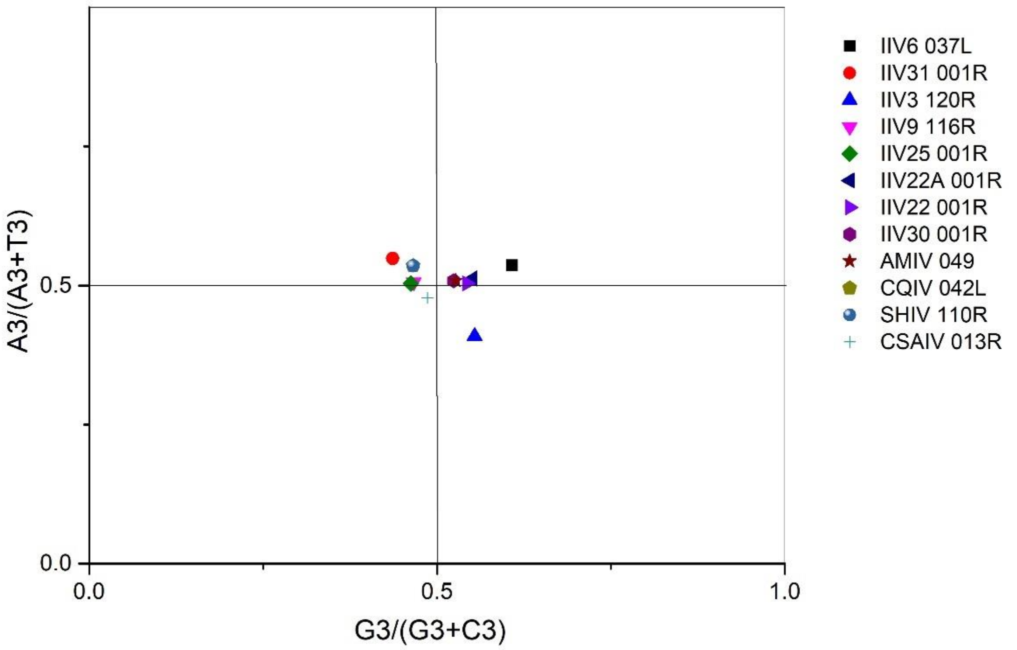

Figure 5. Parity rule 2 (PR2) plot analysis. The PR2 plot was created for the IIVs DNA pol genes.

Table 4. The correlation among the codon compositions (A3s, T3s, G3s, C3s, and GC3s), ENC values, nucleotide compositions (A, T, G, C, and GC), first and second axis values, Gravy values, and Aroma values of DNA pol genes in IIVs.

	A	**C**	**G**	**T**	**GC**	**1st axis**	**2nd axis**	**Gravy**	**Aroma**
T3s	0.91712*	-0.92281*	0.96309*	-0.98589*	-0.98596*	-0.99298*	0.05965	0.18596	0.46667
C3s	-0.98061*	0.98596*	-0.87171*	0.87988*	0.94386*	-0.99298*	0.05965	0.18596	0.46667
A3s	0.99472*	-1	0.86468*	-0.88695*	-0.95789*	-0.95088*	0.08772	0.02456	0.3193
G3s	-0.90301*	0.90877*	-0.95255*	0.98589*	0.97193*	0.98596*	-0.04561	-0.21404	-0.47368
ENC	-0.87479*	0.8807*	-0.77329*	0.77387*	0.8386*	0.84561*	-0.29825	-0.11579	-0.34035
GC3s	-0.9665*	0.97193*	-0.92795*	0.94349*	0.98596*	0.99298*	-0.05965	-0.0807	-0.3614
1st axis	-0.94534*	0.95088*	-0.94903*	0.97176*	0.99298*	1	-0.0386	-0.12281	-0.42456
2nd axis	0.0776	-0.08772	0.25308	-0.06007	-0.04561	-0.0386	1	0.71228*	0.52281
Gravy	-0.00705	-0.02456	0.38313	-0.25796	-0.10877	-0.12281	0.71228*	1	0.63509*
Aroma	0.33863	-0.3193	0.62566*	-0.56185	-0.45965	-0.42456	0.52281	0.63509*	1

The values in the table represent correlation coefficient "r" values derived from a correlation analysis. *P-value< 0.05.

4. Discussion

Invertebrate iridescent viruses (IIVs) are noteworthy for their ability to infect a wide array of invertebrate hosts, leading to diseases that vary in intensity from mild or asymptomatic to severe and potentially lethal (İnce et al., 2018). The genome of IIVs encodes DNA pol protein, which is central to the replication of viral genomes (Jakob et al., 2001). Understanding the specific characteristics of viral DNA pol is essential for developing antiviral therapies and gaining insights into the mechanisms of viral replication. Different viruses have evolved unique strategies for genome replication, and studying their DNA pol is crucial for advancing our knowledge of viral biology and developing targeted interventions. Nevertheless, extensive investigations into the DNA pol genes of invertebrate iridescent viruses remain limited (Nalçacioğlu et al., 2007, 2003).

To enhance our comprehension of the features associated with the DNA pol genes of invertebrate iridescent viruses and to unveil additional details about these viruses, we conducted an analysis focusing on codon usage bias. CUB pertains to the uneven utilization of codons in the encoding of amino acids. In a species genome, CUB is primarily influenced by two key factors: mutation pressure and natural selection (Iriarte et al., 2021; Sharp et al., 1993). Research on the nucleotide composition characteristics and CUB in pathogens is presently restricted. It is recognized that comprehending the patterns of codon usage and the factors influencing them is essential for gaining insights into the genetic evolution of pathogens, including bacteria and viruses (Li et al., 2022; Lu et al., 2023; Wang et al., 2018).

It is widely acknowledged that a genome with an AT-rich content typically harbors codons that end with A and T. Conversely, a genome characterized by a GC-rich content tends to possess codons that terminate with G and C (Ata et al., 2021; Zhong et al., 2007). This discovery echoes earlier studies on the enrichment of A and T in cowpea mild mottle virus (Yang et al., 2022). Nonetheless, the biological implications of this phenomenon remain ambiguous. Hence, it is crucial to investigate the reasons behind the notable increase in A content and simultaneous decrease in C content within the viral genomes (van Hemert and Berkhout, 2016). In our nucleotide composition analysis, we observed a higher occurrence of codons ending with A and T in the IIVs DNA pol genes. Generally, the selection of preferred codons appears to be largely affected by compositional constraints, specifically favoring A and T in the DNA pol coding regions of IIVs. These findings also corroborate the influence of mutation pressure, aligning with previous research (Karumathil et al., 2018; Zhang et al., 2011)

The expression level of a gene is inversely correlated with its ENC value. Generally, if the ENC value is greater than 35, it indicates a genomic composition that is relatively stable and conserved (Rahman et al., 2022; Tyagi et al., 2017). The analysis revealed that the mean ENC value for DNA pol genes is 43.34, signifying a slight bias and a generally consistent codon usage in these genes. Sacrificing translation efficiency and fidelity, the preservation of a slight CUB could serve deliberate aims. For example, it might effectively dampen the expression of genes that trigger innate immunity, thus diminishing competition between the virus and the host and promoting efficient dissemination within the host (Nguyen et al., 2021). Our findings align with those observed in both DNA and RNA viruses, demonstrating a low codon usage bias (Bera et al., 2017; Wang et al., 2023). This likely suggests that the viruses efficiently replicate within hosts, thereby reducing competition with host genes.

The prevailing belief is that the codon usage pattern is primarily affected by mutational pressure and natural selection (Nguyen et al., 2021). However, the specific factors affecting the codon usage pattern of IIVs DNA pol genes remain unclear. We utilized an ENC-GC3s plot analysis to evaluate the factors acting on the codon usage pattern of IIVs DNA pol genes. Our findings illustrated that the data points representing the ENC values for each DNA pol gene were situated below or in close proximity to the anticipated curve. This implies that the codon usage pattern in IIVs DNA pol genes is affected by both mutational bias and translational selection, align with findings from previous research (Feng et al., 2022; Sun et al., 2020; Lu et al., 2023).

If mutation pressure exclusively limits the formation of codon usage patterns, the presence of nucleotides A and T in the 3rd synonymous codon position should be equivalent to that of C and G (Cho et al., 2019; Fu et al., 2023). The PR2 analysis indicated a greater abundance of A and T nucleotides compared to G and C at the third codon position. This finding indicates the substantial impact of natural selection on the formation of codon usage patterns in the DNA pol genes of IIVs. Additionally, neutrality plot analysis demonstrated that natural selection impacts the CUB in DNA pol genes. Similar results have been observed in previous studies as well (Rani et al., 2023; Noor et al., 2023; He et al., 2019).

CAI analysis is employed to assess the optimization of codon usage, gene expression, and the adaptation of viral genes to their hosts (Carbone et al., 2003; Sharp and Li, 1987). A high CAI value indicates a significant level of codon usage bias (Biswas et al., 2019; Khandia et al., 2019). In our research, the CAI analysis revealed that the mean CAI value for DNA pol genes was 0.63, signifying enhanced adaptation of codon usage and expression levels specific to their hosts. The analysis indicated that the codon usage patterns of IIVs DNA pol genes were shaped by the influence of natural selection from their hosts.

In our research, we performed a COA using RSCU to reveal the primary trends in codon usage within DNA pol genes of IIVs and to visually present the findings in a clear and easily understandable manner. The analysis of codon usage unveiled distinct groupings of codons concluding with G-C and A-T on the scatterplot. This suggests a notable inclination towards these specific codon types in the DNA pol genes of IIVs. This inclination is shaped by the nucleotide composition, indicating the presence of varied codon usage patterns and affecting factors within DNA pol genes, which are contingent on the degree of CUB. This observation is in accordance with previous studies (Andargie and Congyi, 2022; Prabha et al., 2012).

The correlation between the nucleotide contents and A3s, T3s, C3s, and G3s provided substantial support for the impact of mutational pressure on the CUB (Wang et al., 2023). Additionally, the notable correlation between ENC values and the entire nucleotide contents affirmed the influence of mutational pressure. The values of the first axes were also strongly correlated with the entire nucleotide contents. All these observations collectively suggest that mutational pressure substantially contributes to the formation of codon usage bias in the DNA pol genes of IIVs. Furthermore, the correlation analysis demonstrated a noteworthy correlation between the aroma value and the G value, indicating the effect of natural selection on the CUB of IIVs DNA pol genes.

5. Conclusions

This investigation delves into the codon usage patterns of DNA pol genes in invertebrate iridescent viruses for the first time, aiming to enhance our understanding of the evolutionary alterations in these viruses. The results from this research revealed a low codon usage bias in the DNA pol genes of invertebrate iridescent viruses. Viruses exhibiting low CUB can employ multiple codons for each amino acid, enhancing their replication efficiency within the host. Moreover, it was shown that, alongside mutational pressure, natural selection also contributed to the shaping of CUB. This investigation not only enhances our comprehension of the variability in the codon usage patterns of DNA pol genes but also aids in identifying the factors that drive the evolution of these genes.

Author Contributions

Y. Aktürk Dizman: Conceptualization, Investigation, Writing – original draft, Software, Formal analysis, Validation, Writing-review & editing.

Conflict of Interest

All the authors declare no conflict of interest.

References

Aktürk Dizman, Y. (2023). Codon usage bias analysis of the gene encoding NAD+-dependent DNA ligase protein of Invertebrate iridescent virus 6. *Archives of Microbiology*, 205(11), 1–19. https://doi.org/10.1007/s00203-023-03688-5

Andargie, M., Congyi, Z. (2022). Genome-wide analysis of codon usage in sesame (*Sesamum indicum* L.). *Heliyon*, 8(1), e08687. https://doi.org/10.1016/j.heliyon.2021.e08687

Ata, G., Wang, H., Bai, H., Yao, X., Tao, S. (2021). Edging on mutational bias, induced natural selection from host and natural reservoirs predominates codon usage evolution in hantaan virus. *Frontiers in Microbiology*, 12, 1–18. https://doi.org/10.3389/fmicb.2021.699788

Bassetto, M., Van Dycke, J., Neyts, J., Brancale, A., Rocha-Pereira, J. (2019). Targeting the viral polymerase of diarrhea-causing viruses as a strategy to develop a single broad-spectrum antiviral therapy. *Viruses*, 11(2), 173.

https://doi.org/10.3390/v11020173

Behura, S.K., Severson, D.W. (2013). Codon usage bias: Causative factors, quantification methods and genome-wide patterns: With emphasis on insect genomes. *Biological Reviews*, 88(1), 49–61. https://doi.org/10.1111/j.1469-185X.2012.00242.x

Bera, B.C., Virmani, N., Kumar, N., Anand, T., Pavulraj, S., Rash, A., Elton, D., Rash, N., Bhatia, S., Sood, R., Singh, R.K., Tripathi, B.N. (2017). Genetic and codon usage bias analyses of polymerase genes of equine influenza virus and its relation to evolution. *BMC Genomics*, 18(1), 652. https://doi.org/10.1186/s12864-017-4063-1

Biswas, K.K., Palchoudhury, S., Chakraborty, P., Bhattacharyya, U.K., Ghosh, D.K., Debnath, P., Ramadugu, C., Keremane, M.L., Khetarpal, R.K., Lee, R.F. (2019). Codon usage bias analysis of *Citrus tristeza* virus: Higher codon adaptation to *Citrus reticulata* host. *Viruses*, 11(4), 1–17. https://doi.org/10.3390/v11040331

Canuti, M., Large, G., Verhoeven, J.T.P., Dufour, S.C. (2022). A novel iridovirus discovered in deep-sea carnivorous sponges. *Viruses*, 14(8), 1595. https://doi.org/10.3390/v14081595

Carbone, A., Zinovyev, A., Képès, F. (2003). Codon adaptation index as a measure of dominating codon bias. *Bioinformatics*, 19(16), 2005–2015. https://doi.org/10.1093/bioinformatics/btg272

Chakraborty, S., Yengkhom, S., Uddin, A. (2020). Analysis of codon usage bias of chloroplast genes in *Oryza* species: Codon usage of chloroplast genes in *Oryza* species. *Planta*, 252(4), 1–20. https://doi.org/10.1007/s00425-020-03470-7

Chen, G., Fang, Y., Yan, Q., Li, P., Wu, L., Feng, G. (2019). The deficiency in nuclear localization signal of *Neodiprion lecontei* nucleopolyhedrovirus DNA polymerase prevents rescue of viral DNA replication and virus production in DNApol-null *Autographa californica* multiple nucleopolyhedrovirus. *Virus Research*, 266, 52–57. https://doi.org/https://doi.org/10.1016/j.virusres.2019.04.005

Chen, L., Liu, T., Yang, D., Nong, X., Xie, Y., Fu, Y., Wu, X., Huang, X., Gu, X., Wang, S., Peng, X., Yang, G. (2013). Analysis of codon usage patterns in *Taenia pisiformis* through annotated transcriptome data. *Biochemical and Biophysical Research Communications*, 430(4), 1344–1348. https://doi.org/10.1016/j.bbrc.2012.12.078

Chen, Y., Xu, Q., Tan, C., Li, X., Chi, X., Cai, B., Yu, Z., Ma, Y., Chen, J.L. (2017). Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Senecavirus A evolution. *Microbial Pathogenesis*, 112, 313–319. https://doi.org/10.1016/j.micpath.2017.09.040

Cho, M., Kim, H., Son, H.S. (2019). Codon usage patterns of LT-AG genes in polyomaviruses from different host species. *Virology Journal*, 16(1), 1–24. https://doi.org/10.1186/s12985-019-1245-2

Coen, D.M., Lawler, J.L., Abraham, J. (2021). Herpesvirus DNA polymerase: Structures, functions, and mechanisms. *Enzymes*, 50, 133–178. https://doi.org/https://doi.org/10.1016/bs.enz.2021.09.003

Delhon, G., Tulman, E.R., Afonso, C.L., Lu, Z., Becnel, J.J., Moser, B.A., Kutish, G.F., Rock, D.L. (2006). Genome of invertebrate iridescent virus type 3 (mosquito iridescent virus). *Journal of Virology*, 80(17), 8439–8449. https://doi.org/10.1128/jvi.00464-06

Dizman, Y.A., Muratoglu, H., Sandalli, C., Nalcacioglu, R., Demirbag, Z. (2016). *Chilo* iridescent virus (CIV) ORF 012L encodes a protein with both exonuclease and endonuclease functions. *Archives of Virology*, 161(11), 3029–3037. https://doi.org/10.1007/s00705-016-3007-4

Eaton, H.E., Metcalf, J., Penny, E., Tcherepanov, V., Upton, C., Brunetti, C.R. (2007). Comparative genomic analysis of the family *Iridoviridae*: Re-annotating and defining the core set of iridovirus genes. *Virology Journal*, 4(11). https://doi.org/10.1186/1743-422X-4-11

Eaton, H.E., Ring, B.A., Brunetti, C.R. (2010). The genomic diversity and phylogenetic relationship in the family *Iridoviridae*. *Viruses*, 2(7), 1458–1475. https://doi.org/10.3390/v2071458

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

Fu, Y., Liang, F., Li, C., Warren, A., Shin, M.K., Li, L. (2023). Codon usage bias analysis in macronuclear genomes of ciliated protozoa. *Microorganisms*, 11(7), 1833. https://doi.org/10.3390/microorganisms11071833

Garro, H.A., Pungitore, C.R. (2019). DNA related enzymes as molecular targets for antiviral and antitumoral chemotherapy. a natural overview of the current perspectives. *Current Drug Targets*, 20(1), 70–80. https://doi.org/10.2174/1389450119666180426103558

He, W., Wang, N., Tan, J., Wang, R., Yang, Y., Li, G., Guan, H., Zheng, Y., Shi, X., Ye, R., Su, S., Zhou, J. (2019). Comprehensive codon usage analysis of porcine deltacoronavirus. *Molecular Phylogenetics and Evolution*. 141, 106618. https://doi.org/10.1016/j.ympev.2019.106618

Hu, Z., Kuritzkes, D.R. (2014). Altered viral fitness and drug susceptibility in HIV-1 carrying mutations that confer resistance to nonnucleoside reverse transcriptase and integrase strand transfer inhibitors. *Journal of Virology*, 88(16):9268-76. https://doi.org/10.1128/jvi.00695-14

İnce, İ.A., Özcan, O., Ilter-Akulke, A.Z., Scully, E.D., Özgen, A. (2018). Invertebrate iridoviruses: A glance over the last decade. *Viruses*, 10(4), 1–25. https://doi.org/10.3390/v10040161

Iriarte, A., Lamolle, G., Musto, H. (2021). Codon usage bias: An endless tale. *Journal of Molecular Evolution*, 89(9-10), 589–593. https://doi.org/10.1007/s00239-021-10027-z

Jakob, N.J., Darai, G. (2002). Molecular anatomy of *Chilo* iridescent virus genome and the evolution of viral genes. *Virus Genes*, 25(3), 299–316. https://doi.org/10.1023/A:1020984210358

Jakob, N.J., Müller, K., Bahr, U., Darai, G. (2001). Analysis of the first complete DNA sequence of an invertebrate iridovirus: Coding strategy of the genome of *Chilo* iridescent virus. *Virology*, 286(1), 182–196. https://doi.org/10.1006/viro.2001.0963

Karumathil, S., Raveendran, N.T., Ganesh, D., Kumar NS, S., Nair, R.R., Dirisala, V.R. (2018). Evolution of synonymous codon usage bias in west african and central african strains of monkeypox virus. *Evolutionary Bioinformatics*, 14. https://doi.org/10.1177/1176934318761368

Khandia, R., Singhal, S., Kumar, U., Ansari, A., Tiwari, R., Dhama, K., Das, J., Munjal, A.O., Singh, R.K. (2019). Analysis of nipah virus codon usage and adaptation to hosts. *Frontiers in Microbiology*, 10, 1–18. https://doi.org/10.3389/fmicb.2019.00886

Komar, A.A. (2016). The Yin and Yang of codon usage. *Human Molecular Genetics*, 25, 77–85. https://doi.org/https://doi.org/10.1093/hmg/ddw207

Li, G., Zhang, L., Xue, P. (2022). Codon usage divergence of important functional genes in *Mycobacterium tuberculosis*. *International Journal of Biological Macromolecules*, 209(2), 1197–1204. https://doi.org/10.1016/j.ijbiomac.2022.04.112

Lu, M., Wan, W., Li, Y., Li, H., Sun, B., Yu, K., Zhao, J., Franzo, G., Su, S. (2023). Codon usage bias analysis of the spike protein of human coronavirus 229E and its host adaptability. *International Journal of Biological Macromolecules*, 253, 127319. https://doi.org/10.1016/j.ijbiomac.2023.127319

Mesplède, T., Quashie, P.K., Osman, N., Han, Y., Singhroy, D.N., Lie, Y., Petropoulos, C.J., Huang, W., Wainberg, M.A. (2013). Viral fitness cost prevents HIV-1 from evading dolutegravir drug pressure. *Retrovirology*, 22(10), 22. https://doi.org/10.1186/1742-4690-10-22

Nalçacioğlu, R., Ince, I.A., Vlak, J.M., Demirbağ, Z., van Oers, M.M. (2007). The *Chilo* iridescent virus DNA polymerase promoter contains an essential AAAAT motif. *Journal of General Virology*, 88(9), 2488–2494. https://doi.org/10.1099/vir.0.82947-0

Nalçacioğlu, R., Marks, H., Vlak, J.M., Demirbaĝ, Z., Van Oers, M.M. (2003). Promoter analysis of the *Chilo* iridescent virus DNA polymerase and major capsid protein genes. *Virology*, 317(2), 321–329. https://doi.org/10.1016/j.virol.2003.08.007

Nambou, K., Anakpa, M. (2020). Deciphering the co-adaptation of codon usage between respiratory coronaviruses and their human host uncovers candidate therapeutics for COVID-19. *Infection, Genetics and Evolution*. 85, 104471. https://doi.org/10.1016/j.meegid.2020.104471

Nguyen, T.H., Wang, D., Rahman, S.U., Bai, H., Yao, X., Chen, D., Tao, S. (2021). Analysis of codon usage patterns and influencing factors in rice tungro bacilliform virus. *Infection, Genetics and Evolution*, 90. https://doi.org/10.1016/j.meegid.2021.104750

Noor, F., Ashfaq, U.A., Bakar, A., Qasim, M., Masoud, M.S., Alshammari, A., Alharbi, M., Riaz, M.S. (2023). Identification and characterization of codon usage pattern and influencing factors in HFRS-causing hantaviruses. Frontiers in Immunology. 14, 1131647. https://doi.org/10.3389/fimmu.2023.1131647

Palanisamy, N., Osman, N., Ohnona, F., Xu, H.T., Brenner, B., Mesplède, T., Wainberg, M.A. (2017). Does antiretroviral treatment change HIV-1 codon usage patterns in its genes: a preliminary bioinformatics study, *AIDS Research and Therapy*, 14(1), 2. https://doi.org/10.1186/s12981-016-0130-y

Prabha, R., Singh, D.P., Gupta, S.K., Farooqi, S., Rai, A. (2012). Synonymous codon usage in *Thermosynechococcus elongatus* (cyanobacteria) identifies the factors shaping codon usage variation. *Bioinformation*, 8(13), 622–628. https://doi.org/10.6026/97320630008622

Puigbò, P., Bravo, I.G., Garcia-Vallve, S. (2008). CAIcal: A combined set of tools to assess codon usage adaptation. *Biology Direct*, 3, 38. https://doi.org/10.1186/1745-6150-3-38

Rahman, S.U., Rehman, H.U., Rahman, I.U., Rauf, A., Alshammari, A., Alharbi, M., Haq, N.U., Suleria, H.A.R., Raza, S.H.A. (2022). Analysis of codon usage bias of lumpy skin disease virus causing livestock infection. *Frontiers in Veterinary Science*, 9, 1071097. https://doi.org/10.3389/fvets.2022.1071097

Rani, S., Mamathashree, M.N., Bharthi, I.U., Patil, S.S., Krishnamoorthy, P., Shueb, M., Pandey, R.K., Suresh, K.P. (2023). Comprehensive examination on codon usage bias pattern of the Bovine Ephemeral fever virus. *Journal of Biomolecular Structure and Dynamics*. 1, 11. https://doi.org/10.1080/07391102.2023.2258220

Rao, Y., Wang, Z., Chai, X., Nie, Q., Zhang, X. (2014). Hydrophobicity and aromaticity are primary factors shaping variation in amino acid usage of chicken proteome. *PLoS One*, 9(10), 1–10. https://doi.org/10.1371/journal.pone.0110381

Schnitzler, P., Soltau, J.B., Fischer, M., Reisner, H., Scholz, J., Delius, H., Darai, G. (1987). Molecular cloning and physical mapping of the genome of insect iridescent virus type 6: Further evidence for circular permutation of the viral genome. *Virology*, 160(1), 66–74. https://doi.org/10.1016/0042-6822(87)90045-6

Sharp, P.M., Li, W.H. (1987). The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, 15(3), 1281–1295. https://doi.org/10.1093/nar/15.3.1281

Sharp, P.M., Li, W.H. (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. *Journal of Molecular Evolution*, 24(19), 28–38. https://doi.org/10.1007/BF02099948

Sharp, P.M., Stenico, M., Peden, J.F., Lloyd, A.T. (1993). Codon usage: mutational bias, translational selection, or both? *Biochemical Society Transactions*, 21(4), 835–841. https://doi.org/10.1042/bst0210835

Smith, R.D. (2022). Enhanced effective codon numbers to understand codon usage bias. *Biosystems*, 220. https://doi.org/https://doi.org/10.1016 /j.biosystems.2022.104734

Sueoka, N. (1999). Two aspects of DNA base composition: G+C content and translation-coupled deviation from intra-strand rule of A = T and G = C. *Journal of Molecular Evolution*, 49(1), 53–58. https://doi.org/10.1007/PL00006534

Sofia, M.J., Chang, W., Furman, P.A., Mosley, R.T., Ross, B.S. (2012). Nucleoside, nucleotide, and non-nucleoside inhibitors of Hepatitis C virus NS5B RNA-dependent RNA-polymerase. *Journal of Medicinal Chemistry*, 55, 2481–2531. https://doi.org/10.1021/jm201384j

Sun, J., Zhao, W., Wang, R., Zhang, W., Li, G., Lu, M., Shao, Y., Yang, Y., Wang, N., Gao, Q., Su, S. (2020). Analysis of the codon usage pattern of HA and NA genes of H7N9 influenza a virus. *International Journal of Molecular Sciences*. 21(19), 7129. https://doi.org/10.3390/ijms21197129

Suzuki, H., Brown, C.J., Forney, L.J., Top, E.M. (2008). Comparison of correspondence analysis methods for synonymous codon usage in bacteria. *DNA Research*, 15(6), 357–365. https://doi.org/10.1093/dnares/dsn028

Tian, H. feng, Hu, Q. mu, Xiao, H. bing, Zeng, L. bing, Meng, Y., Li, Z. (2020). Genetic and codon usage bias analyses of major capsid protein gene in Ranavirus. *Infection, Genetics and Evolution*, 84(8), 1–13. https://doi.org/10.1016/j.meegid.2020.104379

Tyagi, A., Kumar, B.T.N., Singh, N.K. (2017). Genome dynamics and evolution of codon usage patterns in shrimp viruses. *Archives of Virology*, 162(10), 3137–3142. https://doi.org/10.1007/s00705-017-3445-7

Tyagi, A., Nagar, V. (2022). Genome dynamics, codon usage patterns and influencing factors in *Aeromonas hydrophila* phages. *Virus Research*, 320, 198900. https://doi.org/10.1016/j.virusres.2022.198900

van Hemert, F., Berkhout, B. (2016). Nucleotide composition of the Zika virus RNA genome and its codon usage. *Virology Journal*. 13, 95. https://doi.org/10.1186/s12985-016-0551-1

Wang, F., Zhang, N., Zhao, C., Song, Z., Xin, C. (2023). Codon usage bias analysis of mitochondrial protein-coding genes in 12 species of Candida. *Journal of Genetics*, 102(2), 36. https://doi.org/10.1007/s12041-023-01434-w

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

Wang, H., Meng, T., Wei, W. (2018). Analysis of synonymous codon usage bias in helicase gene from *Autographa californica* multiple nucleopolyhedrovirus. *Genes and Genomics*, 40(7), 767–780. https://doi.org/10.1007/s13258-018-0689-x

Williams, T. (2008). Natural invertebrate hosts of iridoviruses (*Iridoviridae*). *Neotrop Entomology*, 37(6), 615–632. https://doi.org/10.1590/S1519-566X2008000600001

Yang, S., Liu, Y., Wu, X., Cheng, X., Wu, X. (2022). Synonymous codon pattern of cowpea mild mottle virus sheds light on its host adaptation and genome evolution. *Pathogens*.11, 419. https://doi.org/10.3390/pathogens11040419

Yannai, A., Katz, S., Hershberg, R. (2018). The codon usage of lowly expressed genes is subject to natural selection. *Genome Biology and Evolution*, 10(5), 1237–1246. https://doi.org/10.1093/gbe/evy084

Yesilyurt, A., Muratoglu, H., Demirbag, Z., Nalcacioglu, R. (2019). *Chilo* iridescent virus encodes two functional metalloproteases. *Archives of Virology*, 164(3), 657–665. https://doi.org/10.1007/s00705-018-4108-z

Zhang, J., Wang, M., Liu, W.Q., Zhou, J.H., Chen, H.T., Ma, L.N., Ding, Y.Z., Gu, Y.X., Liu, Y.S. (2011). Analysis of codon usage and nucleotide composition bias in polioviruses. *Virology Journal*, 8, 146. https://doi.org/10.1186/1743-422X-8-146

Zhang, Y., Shen, Z., Meng, X., Zhang, L., Liu, Z., Liu, M., Zhang, F., Zhao, J. (2022). Codon usage patterns across seven Rosales species. *BMC Plant Biology*, 22(1), 1–10. https://doi.org/10.1186/s12870-022-03450-x

Zhong, J., Li, Y., Zhao, S., Liu, S., Zhang, Z. (2007). Mutation pressure shapes codon usage in the GC-rich genome of foot-and-mouth disease virus. *Virus Genes*, 35(3), 767–776. https://doi.org/10.1007/s11262-007-0159-z

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discovery*. 6,14. https://doi.org/10.1038/s41421-020-0153-3