Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi
Afyon Kocatepe University – Journal of Science and Engineering
https://dergipark.org.tr/tr/pub/akufemubid

Araştırma Makalesi / Research Article
DOI: https://doi.org/10.35414/akufemubid.1388197
AKÜ FEMÜBİD **24** (2024) 031003(492-503) AKU J. Sci. Eng. **24** (2024) 031003(492-503)

A Detailed Analysis of Codon Usages Bias and Affecting Factors in the Topoisomerase II Gene of Invertebrate iridescent virus 6

Makale Bilgisi / Article Info
Alındı/Received: 11.11.2023
Kabul/Accepted: 06.05.2024
Yayımlandı/Published: 27.06.2024

Invertebrate iridescent virüs 6'nın Topoizomeraz II Genindeki Kodon Kullanım Eğiliminin ve Etkileyen Faktörlerin Ayrıntılı Analizi

Yeşim AKTÜRK DİZMAN*

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

Abstract

The topoisomerase II protein (ORF 045L) of invertebrate iridescent virus 6 (IIV6) plays essential roles in managing DNA topology during viral replication and transcription. Considering the importance of the topoisomerase II gene, a comprehensive analysis was conducted to explore the codon usage bias (CUB) of topoisomerase II genes of IIV6 and 9 reference invertebrate iridescent viruses (IIVs). In this research, the findings from the base composition analysis revealed that the IIV6 topoisomerase gene had a high A/T content, with nucleotide A being the most prevalent. The relative synonymous codon usage values for each codon demonstrated the presence of CUB. The effective number of codons (ENC) value for the IIV6 topoisomerase II gene is 34.80, signifying a significant CUB. The ENC plot indicates that all the diverse sequences lie beneath the standard curve, signifying that CUB is influenced not only by mutational pressure but also by other factors, including natural selection. The findings from the neutrality analysis indicate that the codon usage pattern (CUP) is more significantly shaped by natural selection, as evidenced by a regression line slope of 0.1602, compared to the influence of mutation pressure. Furthermore, it has been established that the nucleotide composition and dinucleotide content influence the CUB of the topoisomerase II gene in IIV6. The initial comprehensive analysis of CUB in the IIV6 topoisomerase II gene offers valuable insights into the gene's evolutionary processes.

Keywords: *Invertebrate iridescent virus 6; Topoisomerase II gene; 045L; Codon usage bias*

Öz

Invertebrate iridescent virüs 6 (IIV6)'nın topoizomeraz II proteini (ORF 045L), viral replikasyon ve transkripsiyon sırasında DNA topolojisinin belirlenmesinde önemli roller oynar. Topoizomeraz II geninin önemi göz önünde bulundurularak, IIV6 ve 9 referans invertebrate iridescent virüs (IIVs)'ün topoizomeraz II genlerinin kodon kullanım eğilimini (CUB) araştırmak için kapsamlı bir analiz yapılmıştır. Bu araştırmada, baz bileşimi analizinden elde edilen bulgular, IIV6 topoizomeraz geninin yüksek bir A/T içeriğine sahip olduğunu ve A nükleotidinin en yaygın olduğunu göstermiştir. Her bir kodon için göreli sinonim kodon kullanım değerleri, kodon kullanım eğiliminin varlığını göstermiştir. IIV6 topoizomeraz II geni için etkin kodon sayısı (ENC) değeri 34,80'dir ve bu da önemli bir CUB'ye işaret etmektedir. ENC grafiği, tüm farklı dizilerin standart eğrinin altında yer aldığını ve kodon kullanım eğiliminin sadece mutasyon baskısından değil, doğal seçilim de dahil olmak üzere diğer faktörlerden de etkilendiğini göstermektedir. Nötraliti analizinden elde edilen bulgular, mutasyon baskısının etkisine kıyasla, 0,1602'lik regresyon çizgisi eğiminin de gösterdiği gibi, kodon kullanım modelinin doğal seçilim tarafından daha belirgin bir şekilde şekillendirildiğini göstermektedir. Ayrıca, nükleotid bileşiminin ve dinükleotid içeriğinin IIV6 topoizomeraz II geninin kodon kullanım eğilimini etkilediği tespit edilmiştir. IIV6 topoizomeraz II genindeki kodon kullanım eğiliminin ilk kapsamlı analizi, genin evrimsel süreçleri hakkında değerli bilgiler sunmaktadır.

Anahtar Kelimeler: *Invertebrate iridescent virüs 6; Topoizomeraz II geni; 045L; Kodon kullanım eğilimi*

1. Introduction

Invertebrate iridescent virus 6 (IIV6) is a member of the Iridoviridae family, which is a family of large dsDNA viruses that primarily infect invertebrates, particularly insects and crustaceans, as well as some vertebrates like amphibians and fish (Chinchar et al. 2017). IIV6 is highly pathogenic on important pest insects. Therefore, IIV6 has the potential to function an effective biopesticides (Hernandez et al. 2000, Kleespies et al. 1999). IIV6, like other iridoviruses, is known for its distinctive iridescent appearance, which is a result of its crystalline protein structures (Yan et al. 2009). The IIV6 genome is roughly 212 kilobases long and comprises 215 open reading frames (ORFs), with each ORF having the potential to code for proteins (Eaton et al. 2007). Among these ORFs, one has been identified as a putative topoisomerase II (designated as ORF 045L) based on bioinformatics

*Sorumlu Yazar/Corresponding Author: Yeşim AKTÜRK DİZMAN e-posta/e-mail: yesim.akturk@erdogan.edu.tr

analysis. The IIV6 topoisomerase II protein (ORF 045L) bears a notable resemblance to the topoisomerase II proteins found in other iridoviruses. Viruses are highly diverse, and not all viruses possess a topoisomerase II gene. Some viruses, especially large, double-stranded DNA viruses like iridoviruses, herpesviruses, arboviruses, chlorellaviruses and poxviruses, encode their own topoisomerase II enzymes (Afonso et al. 2006, Coelho and Leitão 2020, Ebert et al. 1990, Fortune et al. 2001, Jakob and Darai 2002).

Topoisomerase II, also known as DNA topoisomerase II, is an essential enzyme involved in the management of DNA topology, which refers to the three-dimensional structure of DNA. It plays a crucial role in various cellular processes, particularly in DNA replication and transcription (Berger 1998). Topoisomerase II is primarily responsible for introducing transient double-stranded breaks in DNA molecules (Champoux 2001). Moreover, the study of viral topoisomerase II has implications for antiviral drug development, as inhibiting this enzyme can disrupt the viral life cycle (Afowowe et al. 2022).

Codon usage pattern significantly impacts viral evolution (Deb et al. 2021). The genetic code exhibits redundancy, as it comprises 64 codons that represent 20 distinct amino acids, resulting in numerous synonymous codons (Spencer and Barral 2012). Nonetheless, patterns of codon usage are not arbitrary, and certain codons are favored over others in various species, which is referred to codon usage bias (CUB) (Parvathy et al. 2022). CUB can be shaped by a variety of elements, such as nucleotide composition, gene length, natural selection and mutational pressure (Nasrullah et al. 2015). Differences in patterns of codon usage exist across viruses, and analyzing codon usage assists in uncovering the evolutionary processes within the genetic composition of viruses and how they adapt to their hosts (Jiang et al. 2022, Kumar et al. 2018).

To date, several genomic sequences of IIVs have been documented. However, research on codon usage bias in iridoviruses has been constrained, and solely a limited set of viral species have been investigated (Aktürk Dizman 2023, Tian et al. 2020, Zhao et al. 2022). There is a notable absence of a comprehensive analysis regarding the CUB in the topoisomerase II gene of IIVs. In this research, diverse techniques were employed to examine the CUP in the topoisomerase II genes of IIV6 and nine other IIVs. The aim was to identify the primary factor influencing the CUB in these viruses. This research will advance our understanding of the molecular evolution of IIVs.

2. Materials and Methods

2.1. Sequence data

The coding sequences of the topoisomerase II gene in ten IIVs were sourced from the GenBank database of the National Center for Biotechnology Information (Int.Ref.1). A list detailing the virus names along with their respective accession numbers is available in Table 1.

2.2. Nucleotide composition analysis

The CodonW software (Int.Ref. 2) was employed for the computation of various parameters, including the occurrences of the nucleotides A, T, G, and C, the occurrences of GC and AT, nucleotide occurrences at the 3rd position of synonymous codons (A3s, G3s, C3s, T3s), and the Gravy and Aroma scores. The CAIcal server. (Int. Ref. 3) was utilized to determine the GC content at the 1st, 2nd, and 3rd positions of every codon, referred to as GC1s, GC2s, and GC3s. In the present study, GC12s was established as the mean value of GC1s and GC2s.

2.3. Relative synonymous codon usage (RSCU)

The RSCU enables the assessment of how synonymous codons are employed for every amino acid (Sharp and Li, 1986). The MEGA 11.0 program was used to determine the RSCU values for the topoisomerase II genes in 10 IIVs (Tamura et al. 2021). An RSCU value of 1 signifies that the synonymous codons are used with equal frequency, while values exceeding 1.0 or falling below 1.0 indicate codons that are more abundant or less abundant, respectively. When RSCU values are greater than 1.6 or less than 0.6, this signifies that the codons are overrepresented or underrepresented, respectively (Zhou et al. 2013). For each synonymous codon, we chose the codon with the highest RSCU and the greatest frequency, designating it as the optimal codon for the IIV6 topoisomerase II gene. We then proceeded to compare these optimal codons with those of the host organism, *Chilo suppressali*s.

2.4. Effective number of codons (ENC)

The ENC value measures the degree of CUB present in a gene, providing insight into the diversity of codons utilized in a sequence. The CodonW software was used to compute the ENC values. It spans from 20 (where each synonymous codon is represented by only one codon) to 61 (when all used codons have the same frequencies for encoding amino acids). Lower ENC values signify a strong preference for particular codons, while more elevated values suggest a more even utilization of synonymous codons. The ENC value of 35 or less is indicative of a remarkably notable CUB (Wright 1990, Comeron and Aguadé 1998).

Table 1. Details about the topoisomerase II genes of IIVs utilized in the examination of codon utilization

	Gene	Virus name	Host	Length (bp)
NC_003038.1	*IIV6 045L*	Invertebrate iridescent virus 6 (IIV6)	*Chilo suppressalis*	3399
NC_024451.1	*IIV31 092R*	Invertebrate iridescent virus 31 (IIV31)	*Armadillidium vulgare*	3474
NC_008187.1	*IIV3 086L*	Invertebrate iridescent virus 3 (IIV3)	*Ochlerotatus taeniorhynchus*	3342
NC_023613.1	*IIV25 151R*	Invertebrate iridescent virus 25 (IIV25)	*Simulium vittatum*	3411
NC_021901.1	*IIV22 141R*	Invertebrate iridescent virus 22 (IIV22)	*Simulium vittatum*	3396
NC_023615.1	*IIV22A 147R*	Invertebrate iridescent virus 22A (IIV22A)	*Simulium vittatum*	3396
NC_023611.1	*IIV30 149R*	Invertebrate iridescent virus 30 (IIV30)	*Helicoverpa zea*	3387
NC_015780.1	*IIV9 089R*	Invertebrate iridescent virus 9 (IIV9)	*Oxycanus dirempta*	3408
NC_040612.1	*CQIV 075L*	Cherax quadricarinatus iridovirus (CQIV)	*Litopenaeus vannamei*	3243
ON887238.1	*CSAIV 135L*	Carnivorous sponge associated iridovirus strain MJ4 (CSAIV)	*Chondrocladia grandis*	3243

2.5. ENC-GC3s plot analysis

ENC-GC3s plot is generally employed to uncover the factors that influence CUB (Wright 1990). The expected ENC value was computed by the formula (1) provided below (Hartl et al. 1994):

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

When a data point falls close to the standard curve or directly on it, this suggests that CUB is mainly impacted by mutation pressure. Conversely, if the data point is situated below the standard curve and at a distance from it, it implies that additional factors, particularly natural selection, remarkably contribute to shaping CUB (Liu 2013).

2.6. Neutrality plot analysis

A graph was created using GC12s and GC3s values, and a regression curve was computed to evaluate the impact of natural selection and mutational pressure on codon usage (Sueoka 1988). Each data point represents a distinct sequence that has been selected for analysis. When the slope of the equation is close to 1, it suggests a strong correlation between GC12s and GC3s, with mutation pressure being the prevailing factor. In contrast, if the slope approaches either the x-axis or the y-axis, it signifies that natural selection is the predominant influencing factor (Wu et al. 2020).

2.7. Dinucleotide relative abundance analysis

Dinucleotide relative frequencies in the topoisomerase II genes of IIVs were determined using compseq software (Int. Ref. 4). The occurrences of 16 dinucleotides were determined as follows(2):

$$\rho xy = \frac{fxy}{fyfx} \quad (2)$$

Here, the frequency of nucleotide y is denoted as fy, and the frequency of nucleotide x is denoted as fx. fyfx signifies the expected frequency of the dinucleotide xy, and fxy signifies the observed frequency of the dinucleotide xy. When the value of ρxy falls below 0.78 or exceeds 1.23, it indicates that the dinucleotides are considered underrepresented or overrepresented, respectively (Kariin and Burge 1995).

2.8. Correspondence analysis (COA)

Codon usage bias can vary between different genes. Hence, the COA was employed to assess the interrelationship and variability in codon usage among the topoisomerase II genes of IIVs (Suzuki et al. 2008). The RSCU values for 59 synonymous codons were graphed on a plot with two axes, namely, axis-1 and axis-2. To carry out the COA, the CodonW software was utilized.

2.9. Codon adaptation index analysis

An analysis of the Codon Adaptation Index (CAI) was performed to anticipate the ability of the topoisomerase II genes in IIVs to adapt to their individual host organisms. The CAI values of the topoisomerase II genes in IIVs were evaluated utilizing the CAIcal server. (Int. Ref. 5) (Puigbò et al. 2008). The host reference datasets were acquired from the CoCoPUTs Database (Int. Ref. 6). The CAI values, based on the reference codon usage pattern, fall within a range of 0 to 1, and elevated CAI values signify a more effective viral adaptation to the specific host.

2.10. Correlation analysis

A Spearman's rank correlation analysis was performed to assess the relationships between nucleotide composition,

ENC, Aromo, Gravy, and the initial two axes of COA. Statistical analyses were executed using OriginPro 9.0.

3. Results

3.1. Nucleotide composition of the IIVs topoisomerase genes

To explore the potential effect of nucleotide constraints in codon usage, the nucleotide composition of the topoisomerase II genes in IIVs was first assessed (Table 2).

The average percentages of A, G, C, and T were found to be 36.62%, 18.56%, 15.06%, and 29.08%, respectively. Likewise, in the IIV6 topoisomerase II gene, the nucleotides A and T were the most prevalent, accounting for 40.90% and 30.00% of the composition, followed by G at 16.50% and C at 12.60%. Moreover, the codon composition at the 3rd position, comprising A3s, T3s, G3s, and C3s, were computed, and they exhibited average values of 47.27%, 47.13%, 18.19%, and 19.78%, respectively, and this pattern remained consistent within the topoisomerase II gene of IIV6. Besides, the mean AT content and the mean GC content were found to be 66.37% and 33.63%, respectively. In addition, the average AT3 content was identified at 70.28%, and the average GC3 content was calculated as 27.96%. Similar results were observed in the topoisomerase II gene of IIV6. These results indicate that the topoisomerase II genes in IIVs exhibit a high AT content and a tendency for A/T nucleotides at the 3rd position of codons.

Table 2. Analysis of nucleotide content of IIVs topoisomerase II genes

IIVs topoisomerase II genes	**A%**	**C%**	**T%**	**G%**	**A3s%**	**C3s%**	**T3s%**	**G3s%**	**AT%**	**AT3%**	**GC%**	**GC12%**	**GC3s%**	**ENC**	**CAI**
IIV6 045L	40.90	12.60	30.00	16.50	61.15	11.41	52.35	8.31	70.90	82.70	29.10	35.00	14.30	34.83	0.84
IIV31 092R	37.30	16.20	26.10	19.90	48.9	24.68	39.75	18.73	63.90	66.00	36.10	37.25	31.90	48.07	0.70
IIV22 141R	34.50	10.30	34.50	14.60	55.87	6.69	63.44	7.01	75.00	88.60	25.00	31.75	9.70	34.72	0.50
IIV22A 147R	40.50	10.20	34.50	14.70	56.49	6.58	63.28	6.73	75.10	88.90	24.90	31.85	9.50	34.57	0.50
IIV30 149R	40.40	10.60	34.20	14.70	55.92	7.99	62.27	7.06	74.60	87.50	25.40	31.85	10.80	35.74	0.74
IIV25 151R	39.90	12.00	32.20	15.90	54.83	10.52	56.07	10.00	72.10	83.60	27.90	33.65	14.80	40.70	0.50
IIV9 089R	39.60	12.00	32.60	15.80	55.49	9.70	56.35	9.65	72.20	84.40	27.80	33.90	14.00	39.99	0.84
IIV3 086L	26.50	25.20	21.90	26.40	17.97	44.62	22.08	46.93	48.40	30.70	51.60	42.80	68.50	46.72	0.54
CSAIV 135L	28.60	25.20	19.40	26.80	16.54	50.97	19.83	44.71	48.10	27.10	51.90	41.45	71.60	48.18	0.54
CQIV 075L	38.00	16.30	25.40	20.30	49.56	24.63	35.84	22.73	63.40	63.30	36.60	36.60	34.50	44.28	0.58
Mean	36.62	15.06	29.08	18.56	47.27	19.78	47.13	18.19	66.37	70.28	33.63	35.61	27.96	40.78	0.63
SD	5.174	5.768	5.534	4.698	16.200	16.260	16.677	15.521	10.421	23.589	10.421	3.941	23.874	5.698	0.138

3.2. RSCU analysis

The RSCU analysis identified preferences for certain synonymous codons that encode the same amino acid. The findings indicated that in the topoisomerase II gene of IIV6, the majority of the favored codons are those that end with A or T (Table 3). Out of the eighteen most prevalent codons (bold, in Table 3), ten concluded with A (TTA, TCA, CCA, ACA, GCA, CAA, AAA, GAA, AGA, and GGA), while eight terminated with T (TTT, ATT, GTT, TAT, CAT, AAT, GAT, and TGT). Of these favored codons, thirteen had RSCU values exceeding 1.6, while three had RSCU values below 0.6. This suggests a notable bias in the usage of synonymous codons. A comparison was made between the CUB of the topoisomerase II gene in IIV6 and those of its host to explore the possible influence of the host species on the CUB of this gene (Table 3). This analysis demonstrated that 54 synonymous codons in the topoisomerase II gene of IIV6 were similar to those found in *Chilo suppressalis*.

3.3. Codon usage bias of IIVs topoisomerase genes

To unravel the factors affecting codon preference in the topoisomerase II genes of IIVs, an ENC-GC3s plot was generated. As depicted in Figure 1, some data points closely align with the theoretical fitting curve, while others are situated below this curve. These findings signify the contribution of both mutation pressure and natural selection in molding the codon usage pattern of

Table 3. The RSCU patterns of the topoisomerase II gene of the IIV6

		RSCU				RSCU	
AA	**Codon**	***IIV6 045L***	***C. Suppressalis***	**AA**	**Codon**	***IIV6 045L***	***C. suppressalis***
Phe (F)	TTT	**1.87**	**1.71**	Ala (A)	GCT	1.10	**2.07**
	TTC	0.13	0.29		GCC	0.55	0.40
Leu (L)	TTA	**3.07**	**3.54**		GCA	**2.21**	1.49
	TTG	0.64	0.49		GCG	0.14	0.04
	CTT	1.64	0.85	Tyr (Y)	TAT	**1.47**	**1.75**
	CTC	0.07	0.23		TAC	0.53	0.25
	CTA	0.50	0.77	His (H)	CAT	**1.54**	**1.55**
	CTG	0.07	0.11		CAC	0.46	0.45
Ile (I)	ATT	**1.45**	1.70	Gln (Q)	CAA	**1.88**	**1.66**
	ATC	0.45	0.30		CAG	0.12	0.34
	ATA	1.10	**1.80**	Asn (N)	AAT	**1.44**	**1.68**
Val (V)	GTT	**2.65**	1.52		AAC	0.56	0.32
	GTC	0.18	0.24	Lys (K)	AAA	**1.66**	**1.85**
	GTA	0.94	**2.00**		AAG	0.34	0.15
	GTG	0.24	0.24	Asp (D)	GAT	**1.62**	**1.63**
Ser (S)	TCT	1.95	**2.13**		GAC	0.38	0.37
	TCC	0.08	0.75	Glu (E)	GAA	**1.77**	**1.71**
	TCA	**2.42**	1.83		GAG	0.23	0.29
	TCG	0.08	0.19	Cys (C)	TGT	**1.60**	**1.38**
	AGT	1.01	0.62		TGC	0.40	0.63
	AGC	0.47	0.58	Arg (R)	CGT	0.00	0.42
Pro (P)	CCT	1.43	1.38		CGC	0.00	0.53
	CCC	0.19	1.22		CGA	0.35	**2.84**
	CCA	**2.29**	**1.38**		CGG	0.00	0.21
	CCG	0.10	0.03		AGA	**5.47**	1.41
					AGG	0.18	0.49
Thr (T)	ACT	1.23	**1.48**	Gly (G)	GGT	1.16	1.12
	ACC	0.25	0.98		GGC	0.23	0.12
	ACA	**2.28**	1.39		GGA	**2.38**	**2.44**
	ACG	0.25	0.14		GGG	0.23	0.32

IIVs topoisomerase II genes. Neutrality plots are commonly utilized for assessing the relative effects of mutation and natural selection. The analysis of the neutrality plot revealed a significant correlation between GC12s and GC3s in the topoisomerase II genes of IIVs ($p < 0.05$, $R^2 = 0.9343$) (Figure 2). Nevertheless, the calculated slope of the regression line within the topoisomerase II gene of IIVs was determined to be 0.1602. This indicates that direct mutation pressure contributes to 16.02% of the influence, while natural selection accounts for 83.98% in influencing the CUB. This highlights that, overall, CUB is predominantly shaped by natural selection.

3.5. The dinucleotide composition analysis in IIVs topoisomerase II genes

In order to evaluate whether dinucleotide compositional constraints could impact the CUP of IIVs topoisomerase II genes, the relative frequencies of 16 dinucleotides were determined (Table 4). The outcomes indicated that CC is overrepresented, while CG, GC, and TA are underrepresented, implying that dinucleotide

occurrences are not random. For the IIV6 topoisomerase II gene, the overrepresented dinucleotides were CA (1.277), TG (1.248) and TT (1.271), while the underrepresented dinucleotides were CG (0.341), GC (0.681) and TA (0.707) (Figure 3). CG emerged as the most notably underrepresented dinucleotide, which aligns with the fact that GC nucleotide content is the lowest that none of the eighteen most commonly utilized synonymous codons terminated with G or C. Among codons containing the TA sequence, not all of them were favored for their corresponding amino acids, with the exception of TTA (Leucine) and TAT (Tyrosine). These findings indicate that the CUB in the topoisomerase II gene of IIV6, as well as that of other IIVs topoisomerase II genes, is affected by dinucleotide composition.

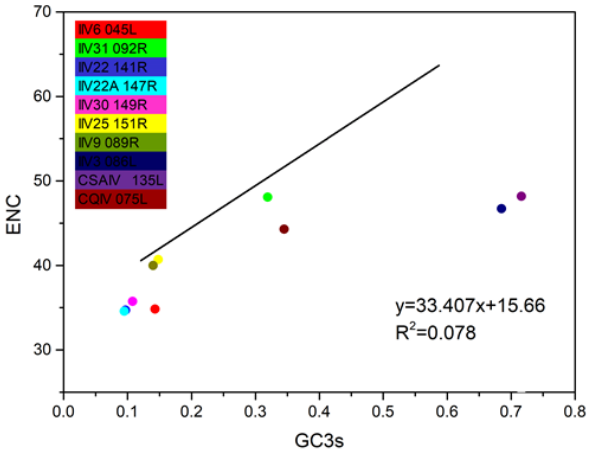

Figure 1. ENC-GC3s plot for IIVs topoisomerase II genes. The plot displays the association between the effective number of codons (ENC) and the GC content at the third position of synonymous codons. The solid black line corresponds to the expected ENC values.

Table 4. The relative frequency of 16 dinucleotides in the topoisomerase II genes

Dinucleotides	Range	Means ± SD
AA	0.994-1.303	1.164 ± 0.080
AC	0.918-1.123	1.007 ± 0.061
AG	0.674-1.018	0.867 ± 0.093
AT	0.763-1.026	0.867 ± 0.068
CA	0.904-1.277	1.064 ± 0.136
CC	0.840-1.737	1.355 ± 0.331
CG	0.311-1.045	0.628 ± 0.284
CT	0.695-1.164	0.990 ± 0.159
GA	0.964-1.193	1.043 ± 0.084
GC	0.574-0.932	0.736 ± 0.106
GG	0.928-1.348	1.194 ± 0.140
GT	0.722-1.077	0.962 ± 0.101
TA	0.486-0.816	0.682 ± 0.145
TC	0.915-1.331	1.043 ± 0.143
TG	1.110-1.318	1.222 ± 0.070
TT	0.904-1.441	1.209 ± 0.149

3.6. Codon adaptation index (CAI) analysis

CAI can provide insights into how well a gene is adapted to its host. CAI values were computed for all codons present within the topoisomerase II genes, employing the codon usage information of eight distinct hosts (Table 1). The CAI value for the IIV6 topoisomerase II gene was 0.84, which notably exceeded the values of others (Table 2). The results of this study reveal a trend to elevated CAI values (≥ 0.5) in the topoisomerase II genes of IIVs, implying a more effective adapted to their specific host organisms.

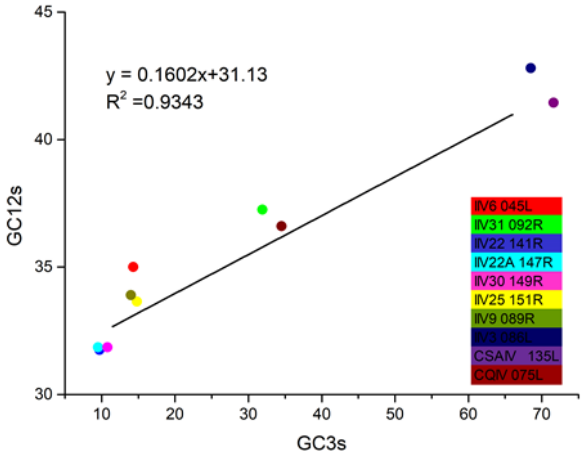

Figure 2. Neutrality plot analysis in IIVs topoisomerase II genes. GC12s denotes the average GC content in the first and second positions of codons (GC1s and GC2s), whereas GC3s signifies the GC content in the third position. The solid line indicates the linear regression relationship between GC12s and GC3s.

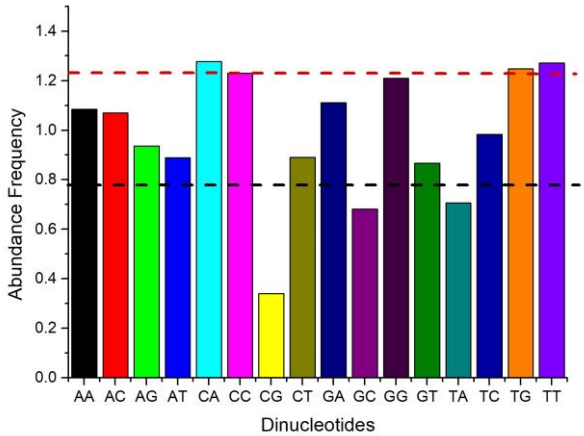

Figure 3. Dinucleotide abundance frequencies in the IIV6 topoisomerase II gene. The lines signify overrepresented values (represented by red dotted lines) and underrepresented values (represented by black dotted lines). The different colored bars depict the frequencies of dinucleotides within the IIV6 topoisomerase II gene.

3.7. Correspondence analysis

Correspondence analysis (COA), a frequently utilized multivariate statistical technique in codon usage analysis, has been employed to examine the main trend of variability within codon usage across genes. COA was carried out utilizing the values of RSCU of 59 sense codons. In line with the COA analysis results, it explained 69.09% of the overall variations, while axis 2 made a contribution of 2.61% to the overall variations. Later on, a scatterplot was created for codons, employing data from the first two axes (Figure 4). The codon scatterplot exhibited prominent groupings that were associated with the nucleotide content in the 3rd codon position.

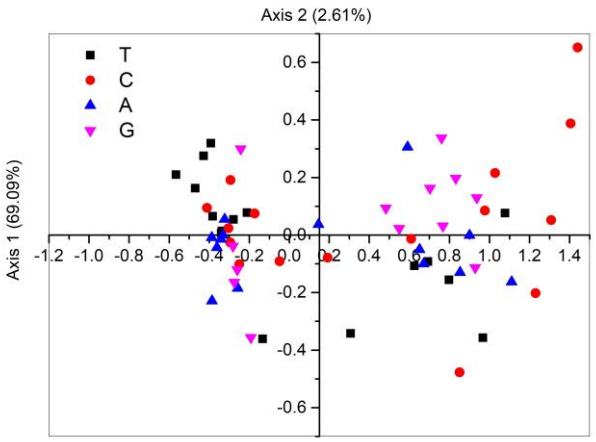

Figure 4. Correspondence analysis of topoisomerase II genes in IIVs. COA was carried out using the 59 synonymous codons' RSCU values. Every symbol in the plot corresponds to distinct codons ending with various bases.

The scatterplot clearly depicted that codons that terminate with G-C and those that end with A-T are organized into diverse and readily distinguishable clusters.

3.8. Correlation analysis of IIVs topoisomerase II genes

To examine CUB in the topoisomerase II genes, the two primary factors, namely natural selection and mutational pressure, were taken into account. To explore the impact of mutation pressure on CUB, we conducted a correlation analysis to examine the relationships between A3s, T3s, G3s, C3s, GC3s, ENC values, and nucleotide contents (Table 5).

A Significant correlation was detected among these variables. Additionally, the correlation analysis indicated that axis 1 positively correlated with ENC, C3s, and GC3s, and negatively correlated with A3s and T3s. These findings suggested that mutational pressure could have a significant impact on shaping CUB of IIV topoisomerase II genes.

Correlation analysis was employed to evaluate the connection between CUB and the Gravy and Aroma scores (Table 5), revealing the influence of natural selection. The findings indicated that the codon compositions did not exhibit a notable correlation with the Gravy and Aroma scores. Only a minor correlation was observed between the Gravy score and axis 2, revealing that natural selection contributes to shaping the CUB in the topoisomerase II genes of IIVs.

Table 5. The correlation among the codon compositions (A3s, T3s, G3s, C3s, and GC3s), the ENC values, nucleotide compositions (A, T, G, C, and GC), the 1st axis values, the 2nd axis values, the Gravy values, and the Aroma values in the topoisomerase II genes of IIVs

	A	C	G	T	GC	1st axis	2nd axis	Gravy	Aroma
T3s	0.50303	-0.98173*	-0.99697*	0.99697*	-0.98788*	-0.93939*	0.4303	0.2	0.31611
C3s	-0.58788	0.98173*	-0.98481*	-0.98481*	0.98788*	0.93939*	-0.38182	-0.22424	-0.29179
A3s	0.85455*	-0.7866*	-0.76596*	0.7842*	-0.79394*	-0.89091*	0.06667	-0.15152	0.50456
G3s	-0.67273*	-0.94833*	0.9301*	0.95733*	0.95152*	1	-0.24848	0.04242	-0.32827
ENC	-0.66061*	0.87807*	0.86323*	-0.88754*	0.89091*	0.92727*	-0.30909	-0.09091	-0.49848
GC3s	-0.61212	0.97563*	0.96657*	-0.98481*	0.98788*	0.97576*	-0.3697	-0.11515	-0.38906
2nd axis	-0.27273	-0.36586	-0.43769	0.41945	-0.40606	-0.24848	1	0.73333*	0.56535

4. Discussion

CUB pertains to the uneven utilization of codons in encoding amino acids. While mutation and natural selection are the predominant forces impacting CUB in the genome of a species, other elements like nucleotide composition, evolutionary pressure, and geographical distribution could also exert an impact (Iriarte et al. 2021, Shi et al. 2016). Extensive research has been conducted on the nucleotide content properties and CUB in complex organisms like Drosophila and mammals (Machado et al. 2020, Ouyang et al. 2022). However, comparable studies involving pathogens are relatively scarce. Understanding codon usage patterns and the elements that influence them is considered essential for comprehending the genetic evolution of infectious agents, including bacteria and viruses (Begum and Chakraborty 2022, Sharma et al.

2023). In order to address the knowledge voids in the examination of IIVs, we employed various analytical methods in this research to explore the codon usage patterns and uncover the factors that impact CUB in the topoisomerase genes of IIVs. The general nucleotide composition significantly impacts codon usage preferences (Xu et al. 2017). Based on our analysis, the topoisomerase II gene of IIVs exhibits a highly pronounced AT richness, especially noticeable in the third position of synonymous codons. It has been proposed that in viruses, especially those with AT or GC-rich genomes, there tends to be a correlation with RSCU patterns. For example, codons with a preference for AT-rich or GC-rich compositions typically end with either A and T or G and C, respectively. These observed trends provide evidence of the effect of mutational pressure (Jenkins and Holmes 2003).

The RSCU analysis demonstrated that there was a notable inclination towards codons ending with A and T in the IIV6 topoisomerase II gene. To account for potential host-specific patterns, we computed RSCU values specific to the host. Consequently, we conducted a comparative analysis between the RSCU patterns of the IIV6 topoisomerase II gene and those of its respective host. It was noted that the codon usage patterns specific to the host also displayed a preference for codons that conclude with A and T. Therefore, consistent with the analysis of nucleotide composition, the RSCU analysis provided additional confirmation of the bias toward codons ending with A and T. Hence, it can be inferred that mutational bias was identified to be the major force shaping the codon usage patterns in the topoisomerase II gene of IIV6, indicating that compositional constraints were a contributing factor in the choice of favored codons.

The analysis of ENc values indicates that the CUB in the IIVs topoisomerase II gene is minimal. A potential advantage of this low CUB is the enhancement of gene transcription and translation efficiency for the IIVs topoisomerase II gene. In fact, several viruses exhibit a similarly low codon usage bias, including the Rabies virus, PEDV, and HCV (Khandia et al. 2023, Yu et al. 2021, Zhang et al. 2018). Nevertheless, our research found that the IIV6 topoisomerase II gene displayed a more significant CUB in comparison to the remaining topoisomerase II genes.

Mutational pressure and natural selection are two crucial components that have impacted codon usage patterns, a phenomenon that is widely observed across the genomes of various viruses and is significantly shaped by the evolution of the genome (Tao et al. 2009). To appraise the effectiveness of natural selection and mutation in influencing CUB in viruses, ENC-plot and neutrality analysis were conducted. These results indicate that the CUB sighted in the topoisomerase II genes of IIVs is likely influenced by both mutational pressure and natural selection rather than mutational pressure alone. These results align with previous research (Patil et al. 2021, Nyayanit et al. 2021).

The frequencies of 16 dinucleotides were examined to explore the possible influence of dinucleotides. Among these dinucleotides, the CG content in the IIVs topoisomerase II genes was notably the lowest. Previous research has indicated that in the evolutionary process, many viruses tend to decrease the presence of CG in their genomic elements. This reduction offers advantages related to immune evasion or host adaptation (Molteni et al. 2023, Sharp et al. 2023). Our findings corroborated the hypothesis that a decreased CG content within a viral genome might provide an advantage for adapting to host organisms (Shackelton et al. 2006, Si et al. 2021). TA is another dinucleotide commonly found to be underrepresented in the genomes of both DNA and RNA viruses (Aktürk Dizman 2023, Kumar et al. 2016). This is due to TA dinucleotides being more prone to degradation by cytoplasmic RNase, a mechanism that contributes to the regulation of mRNA turnover in the cell (Odon et al. 2019). This analysis implied that dinucleotide contents have a considerable impact on shaping the CUP in IIVs topoisomerase II genes.

To investigate how natural selection impacted codon usage, using the codon usage of the hosts as reference sets, CAI values were computed. As CAI measures the resemblance between synonymous codon usage in a gene and this of a reference set, elevated CAI values signify a strong codon usage bias, reflecting the dominant impact of natural selection (Carbone et al. 2003). The mean CAI value for IIVs topoisomerase II genes in host organisms is 0.64, indicating a high expression level in these hosts. This implies that the formation of CUB in topoisomerase II genes of IIVs has been influenced by natural selection. These results align with previous studies, emphasizing the significance of natural selection in shaping the CUB of genes present in viruses (He et al. 2019, Tian et al. 2020).

In this investigation, we also carried out COA using RSCU to unveil the principal trends in CUP within the topoisomerase II genes and to present the results visually in a straightforward manner. The COA revealed clear clusters of G/C-ended and A/T-ended codons on the scatterplot, indicating a pronounced preference for these

codon types in IIVs topoisomerase II genes. This preference is impacted by the nucleotide content and indicates that various patterns of codon usage and affecting elements across topoisomerase genes, depending on the level of CUB. This finding aligns with earlier research (Duan et al. 2021, Wei et al. 2014).

Codon usage bias in some viruses is shaped by both mutational pressure and natural selection, both of which are recognized as significant driving factors (Chen 2013). The significance of mutational pressure in molding codon usage bias was additionally supported by the remarkably strong correlations observed between the overall nucleotide compositions and A3s, T3s, G3s, C3s, and GC3s. Additionally, the substantial correlation between the ENC values and the overall nucleotide compositions further reinforced the significance of mutational pressure. The first axis values of COA were also noted to have a significant correlation with the overall nucleotide compositions. These observations validate that mutational pressure is a significant driving force behind the CUB observed in IIVs topoisomerase II genes.

Throughout the process of a virus adapting to a host cell, the codon usage pattern might also be affected by natural selection (Shi et al. 2016). The notable correlation between the Gravy value and axis 2 suggests that the characteristics of viral proteins contribute to the observed differences in codon usage among IIVs topoisomerase II genes. This underscores the significance of natural selection in molding CUB within these genes.

5. Conclusions

Considering all of these findings, our investigations demonstrate that the IIV6 topoisomerase gene possesses a notable codon usage bias, affected by mutational pressure, natural selection, and dinucleotide abundance. To our knowledge, this is the initial extensive study delving into codon usage bias and its determinants in the IIV6 topoisomerase II gene. Our research contributes to a deeper understanding of the evolutionary dynamics of the IIV6 topoisomerase II gene, providing a foundation for future fundamental research in this area.

Declaration of Ethical Standards

The authors declare that they comply with all ethical standards.

Credit Authorship Contribution Statement

Author-1: Conceptualization, investigation, methodology and software, visualization and writing – original draft, supervision and writing – review and editing.

Declaration of Competing Interest

The authors have no conflicts of interest to declare regarding the content of this article.

Data Availability Statement

All data generated or analyzed during this study are included in this published article.

Acknowledgement

I would like to thank Assoc. Prof. Dr. Serdar DİZMAN for his contributions in drawing the graphs in the article and I would also like to thank the referees and editors who contributed to the revision of the article for their support.

6. References

Afonso, C.L., Tulman, E.R., Delhon, G., Lu, Z., Viljoen, G.J., Wallace, D.B., Kutish, G.F. and Rock, D.L., 2006. Genome of Crocodilepox Virus. *Journal of Virology*, **80**, 4978–4991. https://doi.org/10.1128/jvi.80.10.4978-4991.2006

Afowowe, T.O., Sakurai, Y., Urata, S., Zadeh, V.R. and Yasuda, J., 2022. Topoisomerase II as a Novel Antiviral Target against Panarenaviral Diseases. *Viruses*, **15**, 105. https://doi.org/10.3390/v15010105

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**, 352. https://doi.org/10.1007/s00203-023-03688-5

Begum, N. and Chakraborty, S., 2022. Influencing elements of codon usage bias in Birnaviridae and its evolutionary analysis. *Virus Research*, **310**, 198672. https://doi.org/10.1016/j.virusres.2021.198672

Berger, J.M., 1998. Structure of DNA topoisomerases. *Biochimica et Biophysica Acta*, **1400**, 3–18. https://doi.org/10.1016/S0167-4781(98)00124-9

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

Champoux, J.J., 2001. DNA topoisomerases: structure, function, and mechanism. *Annual Review of Biochemistry*, 70, 369–413. https://doi.org/10.1146/annurev.biochem.70.1.36

Chen, Y., 2013. A comparison of synonymous codon usage bias patterns in DNA and RNA virus genomes: Quantifying the relative importance of mutational pressure and natural selection. *BioMed Research International*, **2013**, 406342. https://doi.org/10.1155/2013/406342

Chinchar, V.G., Waltzek, T.B. and Subramaniam, K., 2017. Ranaviruses and other members of the family

Iridoviridae: Their place in the virosphere. *Virology*, **511**, 259–271. https://doi.org/10.1016/j.virol.2017.06.007

Coelho, J. and Leitão, A., 2020. The African swine fever virus (ASFV) topoisomerase ii as a target for viral prevention and control. *Vaccines*, **8**, 1–16. https://doi.org/10.3390/vaccines8020312

Comeron, J.M. and Aguadé, M., 1998. An evaluation of measures of synonymous codon usage bias. *Journal of Molecular Evolution*, **47**, 268–274. https://doi.org/10.1007/PL00006384

Deb, B., Uddin, A. and Chakraborty, S., 2021. Composition, codon usage pattern, protein properties, and influencing factors in the genomes of members of the family Anelloviridae. *Archives of Virology*, **166**, 461–474. https://doi.org/10.1007/s00705-020-04890-2

Duan, H., Zhang, Q., Wang, C., Li, F., Tian, F., Lu, Y., Hu, Y., Yang, H. and Cui, G., 2021. Analysis of codon usage patterns of the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result of major selection constraints. *PeerJ*, 9, e10787. https://doi.org/10.7717/peerj.10787

Eaton, H.E., Metcalf, J., Penny, E., Tcherepanov, V., Upton, C. and 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

Ebert, S.N., Shtrom, S.S. and Muller, M.T., 1990. Topoisomerase II cleavage of herpes simplex virus type 1 DNA in vivo is replication dependent. *Journal of Virology*, **64**, 4059–4066. https://doi.org/10.1128/jvi.64.9.4059-4066.1990

Fortune, J.M., Lavrukhin, O. V., Gurnon, J.R., Van Etten, J.L., Lloyd, R.S. and Osheroff, N., 2001. Topoisomerase II from Chlorella Virus PBCV-1 Has an Exceptionally High DNA Cleavage Activity. *Journal of Biological Chemistry*, **276**, 24401–24408. https://doi.org/10.1074/jbc.M101693200

Hartl, D.L., Moriyama, E.N. and Sawyer, S.A., 1994. Selection intensity for codon bias. *Genetics*, **138**, 227–234. https://doi.org/10.1093/genetics/138.1.227

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

Hernandez, O., Maldonado, G. and Williams, T., 2000. An Epizootic of Patient Iridescent Virus Disease in Multiple Species of Blackflies in Chiapas, Mexico. *Medical and Veterinary Entomology*, **14**, 458-462. https://doi.org/10.1046/j.1365-2915.2000.00258.x

Iriarte, A., Lamolle, G. and Musto, H., 2021. Codon Usage Bias: An Endless Tale. *Journal of Molecular Evolution*, **89**, 589–593. https://doi.org/10.1007/s00239-021-10027-z

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

Jenkins, G.M. and Holmes, E.C., 2003. The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Research*, **92**, 1–7. https://doi.org/10.1016/S0168-1702(02)00309-X

Jiang, L., Zhang, Q., Xiao, S. and Si, F., 2022. Deep decoding of codon usage strategies and host adaption preferences of soybean mosaic virus. *International Journal of Biological Macromolecules*, **222**, 803-817. https://doi.org/10.1016/j.ijbiomac.2022.09.179

Kariin, S. and Burge, C., 1995. Dinucleotide relative abundance extremes: a genomic signature. *Trends in Genetics*, **11**, 283–290. https://doi.org/10.1016/S0168-9525(00)89076-9

Khandia, R., Khan, A.A., Karuvantevida, N., Gurjar, P., Rzhepakovsky, I.V. and Legaz, I., 2023. Insights into Synonymous Codon Usage Bias in Hepatitis C Virus and Its Adaptation to Hosts. *Pathogens*, **12**, 325. https://doi.org/10.3390/pathogens12020325

Kleespies, R.G., Tidona, C.A. and Darai, G., 1999. Characterization of a New Iridovirus Isolated from Crickets and Investigations on the Host Range. *Journal of Invertebrate Pathology*, **73**, 84-90. https://doi.org/10.1006/jipa.1998.4821

Kumar, N., Bera, B.C., Greenbaum, B.D., Bhatia, S., Sood, R., Selvaraj, P., Anand, T., Tripathi, B.N. and Virmani, N., 2016. Revelation of influencing factors in overall codon usage bias of equine influenza viruses. *PLoS One* **11**, 1–26. https://doi.org/10.1371/journal.pone.0154376

Kumar, N., Kulkarni, D.D., Lee, B., Kaushik, R., Bhatia, S., Sood, R., Pateriya, A.K., Bhat, S. and Singh, V.P., 2018. Evolution of codon usage bias in henipaviruses is governed by natural selection and is host-specific. *Viruses*, **10**, 604. https://doi.org/10.3390/v10110604

Liu, X., 2013. A more accurate relationship between "effective number of codons" and GC3s under assumptions of no selection. *Computational Biology and Chemistry*, **42**, 35–39. https://doi.org/10.1016/j.compbiolchem.2012.11.003

Machado, H.E., Lawrie, D.S. and Petrov, D.A., 2020. Pervasive Strong Selection at the Level of Codon Usage Bias in *Drosophila melanogaster*. *Genetics*, **214**, 511–528. https://doi.org/10.1534/genetics.119.302542

Molteni, C., Forni, D., Cagliani, R., Bravo, I.G. and Sironi, M., 2023. Evolution and diversity of nucleotide and dinucleotide composition in poxviruses. *Journal of General Virology*, **104**, 1–19. https://doi.org/10.1099/jgv.0.001897

Nasrullah, I., Butt, A.M., Tahir, S., Idrees, M. and Tong, Y., 2015. Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. *BMC Evolutionary Biology*, **15**, 1–15. https://doi.org/10.1186/s12862-015-0456-4

Nyayanit, D.A., Yadav, P.D., Kharde, R. and Cherian, S., 2021. Natural selection plays an important role in shaping the codon usage of structural genes of the viruses belonging to the coronaviridae family. *Viruse*s, **13**, 3. https://doi.org/10.3390/v13010003

Odon, V., Fros, J.J., Goonawardane, N., Dietrich, I., Ibrahim, A., Alshaikhahmed, K., Nguyen, D. and Simmonds, P., 2019. The role of ZAP and OAS3/RNAseL pathways in the attenuation of an RNA virus with elevated frequencies of CpG and UpA dinucleotides. *Nucleic Acids Research*, **47**, 8061–8083. https://doi.org/10.1093/nar/gkz581

Ouyang, T., Zhong, J., Chai, Z., Wang, J., Zhang, M., Wu, Z. and Xin, J., 2022. Codon Usage Bias and Cluster Analysis of the MMP-2 and MMP-9 Genes in Seven Mammals. *Genetics Research*, **2022**, 2823356. https://doi.org/10.1155/2022/2823356

Parvathy, S.T., Udayasuriyan, V. and Bhadana, V., 2022. Codon usage bias. *Molecular Biology Reports*, **49**, 539–565. https://doi.org/10.1007/s11033-021-06749-4

Patil, S.S., Indrabalan, U.B., Suresh, K.P. and Shome, B.R., 2021. Analysis of codon usage bias of classical swine fever virus. *Veterinary World*, **14**, 1450–1458. https://doi.org/10.14202/vetworld.2021.1450-1458

Puigbò, P., Bravo, I.G. and 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

Shackelton, L.A., Parrish, C.R. and Holmes, E.C., 2006. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *Journal of Molecular Evolution*, **62**, 551–563. https://doi.org/10.1007/s00239-005-0221-1

Sharma, A., Gupta, S. and Paul, K., 2023. Codon usage behavior distinguishes pathogenic Clostridium species from the non-pathogenic species. *Gene*, **873**, 147394. https://doi.org/10.1016/j.gene.2023.147394

Sharp, C., Thompson, B., Nash, T., Diebold, O., Pinto, R., Thorley, L., Lin, Y., Sives, S., Wise, H., Clohisey Hendry, S., Grey, F., Vervelde, L., Simmonds, P., Digard, P. and Gaunt, E., 2023. CpG dinucleotide enrichment in the influenza A virus genome as a live attenuated vaccine development strategy. *PLoS Pathogens*, **19**, e1011357. https://doi.org/10.1371/journal.ppat.1011357

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

Shi, S.L., Jiang, Y.R., Yang, R.S., Wang, Y. and Qin, L., 2016. Codon usage in Alphabaculovirus and Betabaculovirus hosted by the same insect species is weak, selection dominated and exhibits no more similar patterns than expected. *Infection, Genetics and Evolution*, **44**, 412–417. https://doi.org/10.1016/j.meegid.2016.07.042

Si, F., Jiang, L., Yu, R., Wei, W. and Li, Z., 2021. Study on the Characteristic Codon Usage Pattern in Porcine Epidemic Diarrhea Virus Genomes and Its Host Adaptation Phenotype. *Frontiers in Microbiology*, **12**, 1–18. https://doi.org/10.3389/fmicb.2021.738082

Spencer, P.S. and Barral, J.M., 2012. Genetic code redundancy and its influence on the encoded polypeptides. *Computational and Structural Biotechnology Journal*, **1**, e201204006. https://doi.org/10.5936/csbj.201204006

Sueoka, N., 1988. Directional mutation pressure and neutral molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 2653–2657. https://doi.org/10.1073/pnas.85.8.2653

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

Tamura, K., Stecher, G. and Kumar, S., 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, **38**, 3022–3027. https://doi.org/10.1093/molbev/msab120

Tao, P., Dai, L., Luo, M., Tang, F., Tien, P. and Pan, Z., 2009. Analysis of synonymous codon usage in classical

swine fever virus. *Virus Genes*, **38**, 104–112. https://doi.org/10.1007/s11262-008-0296-z

Tian, H. feng, Hu, Q. mu, Xiao, H. bing, Zeng, L. bing, Meng, Y. and Li, Z., 2020. Genetic and codon usage bias analyses of major capsid protein gene in Ranavirus. *Infection, Genetics and Evolution*, **84**, 104379. https://doi.org/10.1016/j.meegid.2020.104379

Wei, L., He, J., Jia, X., Qi, Q., Liang, Z., Zheng, H., Ping, Y., Liu, S. and Sun, J., 2014. Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. *BMC Evolutionary Biology*, **14**, 1–12. https://doi.org/10.1186/s12862-014-0262-4

Wright, F., 1990. The "effective number of codons" used in a gene. *Gene*, **87**, 23–29. https://doi.org/10.1016/0378-1119(90)90491-9

Wu, H., Bao, Z., Mou, C., Chen, Z. and Zhao, J., 2020. Comprehensive Analysis of Codon Usage on Porcine Astrovirus. *Viruses*, **12**, 991. https://doi.org/10.3390/v12090991

Xu, X., Fei, D., Han, H., Liu, H., Zhang, J., Zhou, Y., Xu, C., Wang, H., Cao, H. and Zhang, H., 2017. Comparative characterization analysis of synonymous codon usage bias in classical swine fever virus. *Microbial Pathogenesis*, **107**, 368–371. https://doi.org/10.1016/j.micpath.2017.04.019

Yan, X., Yu, Z., Zhang, P., Battisti, A.J., Holdaway, H.A., Chipman, P.R., Bajaj, C., Bergoin, M., Rossmann, M.G. and Baker, T.S., 2009. The Capsid Proteins of a Large, Icosahedral dsDNA Virus. *Journal of Molecular Biology*, **385**, 1287–1299. https://doi.org/10.1016/j.jmb.2008.11.002

Yu, X., Gao, K., Pi, M., Li, H., Zhong, W., Li, B. and Ning, Z., 2021. Phylogenetic and codon usage analysis for replicase and capsid genes of porcine circovirus 3. *Veterinary Research Communications*, **45**, 353-361. https://doi.org/10.1007/s11259-021-09816-0

Zhang, X., Cai, Y., Zhai, X., Liu, J., Zhao, W., Ji, S., Su, S. and Zhou, J., 2018. Comprehensive analysis of codon usage on rabies virus and other lyssaviruses. *International Journal of Molecular Sciences*, **19**, 2397. https://doi.org/10.3390/ijms19082397

Zhao, R., Gu, C., Zou, X., Zhao, M., Xiao, W., He, M., He, L., Yang, Q., Geng, Y. and Yu, Z., 2022. Comparative genomic analysis reveals new evidence of genus boundary for family Iridoviridae and explores qualified hallmark genes. *Computational and Structural Biotechnology Journal*, **20**, 3493–3502. https://doi.org/10.1016/j.csbj.2022.06.049

Zhou, J., Gao, Z., Zhang, J., Ding, Y., Stipkovits, L., Szathma, S., Pejsak, Z. and Liu, Y., 2013. The analysis of codon bias of foot-and-mouth disease virus and the adaptation of this virus to the hosts. *Infection, Genetics and Evolution*, **14**, 105–110. https://doi.org/10.1016/j.meegid.2012.09.02

Internet References

1-National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov, (1.10.2023)

2-CodonW,http://codonw.sourceforge.net/, (16.10.2023)

3-CAIcal server, http://genomes.urv.es/CAIcal/, (17.10.2023)

4-https://www.bioinformatics.nl/cgi-bin/emboss/compseq, (23.10.2023)

5-RCDI server, http://genomes.urv.cat/CAIcal/RCDI/, (17.10.2023)

6-Codon and Codon Pair Usage Tables, https://dnahive.fda.gov/dna.cgi?cmd=codon_usage &id=537&mode=cocoputs, (15.10.2023)