

RNA editing sites and compositional features in mitochondrial genes of two freshwater fish species *Clarias batrachus* and *Heteropneustes fossilis*

J.C. Das, V. Dey and S. Chakraborty*

Department of Biotechnology, Assam University, Silchar-788 011, India

Received: 01 April 2025

Revised: 09 July 2025

Accepted: 15 October 2025

*Corresponding Author Email: biotechnology700@gmail.com

*ORCID: <https://orcid.org/0009-0005-7853-9553>

Abstract

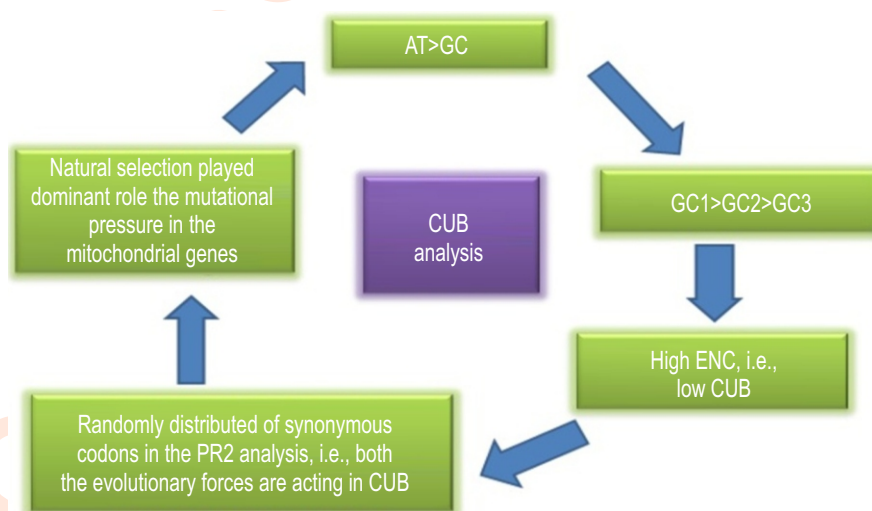
Aim: This study explores the codon usage bias (CUB) in the mitochondrial genes of two economically significant catfish species, *Clarias batrachus* and *Heteropneustes fossilis*, providing valuable insights for genetic and evolutionary research.

Methodology: Effective Number of Codons (ENC), RSCU values, Compositional analysis, Correspondence and PR2 analysis, Neutrality plot, RNA editing etc., were conducted.

Results: A low CUB was observed in both species, with average Effective Number of Codons (ENC) values 43.46 for *C. batrachus* and 44.08 for *H. fossilis*. In *C. batrachus*, 12 codons were over-represented, while 19 were under-represented; in *H. fossilis*, 11 codons were over-represented and 17 under-represented. Compositional analysis showed an AT-rich bias at third codon position in both the species. Correspondence and parity rule 2 (PR2) analyses suggested that mutational pressure and natural selection contribute to CUB. Neutrality plots indicated that natural selection plays a significant role than mutation in shaping CUB in the mitochondrial genes. Additionally, high C to T transitions was observed at RNA editing sites. Positive GRAVY scores pointed to the hydrophobic nature of mitochondrial proteins. Phylogenetic analysis revealed the evolutionary relationships between these species.

Interpretation: This study provides valuable insights into the genetic evolution of these catfish species, with implications for aquaculture and evolutionary biology.

Key words: Codon usage bias, Natural selection, Phylogenetic, RNA editing sites



Introduction

A wide variety of species and habitats can be found in freshwater ecosystems (Collen *et al.*, 2014). *Clarias batrachus* (Magur), a freshwater fish, is able to breathe air and move around on land. It belongs to Family Claridae and inhabits varied low-oxygen settings, burrows into the mudflat, and occasionally “walks” to find appropriate places during summer season (Kumar *et al.*, 2017; Li *et al.*, 2018). *Clarias batrachus* is a highly valued catfish species due to desirable attributes, such as distinct flavour, low fat content, medicinal properties, and mild taste. Its ability to thrive in low-oxygen conditions makes it a commercially viable option for aquaculture (Kushwaha *et al.*, 2015). Indian catfish, *Heteropneustes fossilis* (also called the stinging catfish or Singhi) found in the Indian subcontinent and South-east Asia is regarded as one of the most desirable freshwater species of air-breathing fish (Puvaneswari *et al.*, 2009). *Heteropneustes fossilis*, belongs to Family Heteropneustidae, is distinguished by its elongated body, notably short anal fins, four pairs of barbels, and a venom gland located on the pectoral spines.

Additionally, it possesses a specialized air sac that facilitates aerial respiration, along with a highly depressed head, contributing to its unique morphological adaptations (Diogo *et al.*, 2003; Ratmuangkhwang *et al.*, 2014). Marshes, swamps, ponds, and ditches are primary habitat of these fishes. Since they are durable, tolerant of large stocking rates, and respire using ambient oxygen, these species are perfect for aquaculture (Puvaneswari *et al.*, 2009). They form a significant dietary source of high protein, iron and calcium (Hasan *et al.*, 2022). In vertebrates, the mitochondrial genome is a small, circular DNA molecule, typically spanning 14 to 20 kbp. It encodes 13 protein-coding genes, 2 ribosomal RNA genes, and 22 transfer RNA genes. Additionally, it contains two essential non-coding regions: the L-strand origin of replication (OL) and the control region (Montaña-Lozano *et al.*, 2022). Introns are absent in mitochondrial genome. Nuclear DNA-encoded polymerases help the mtDNA genome in self-replication, and supports the ATP synthase, NADH, and cytochrome oxidase respiratory systems in cells (Uddin *et al.*, 2015). Codons that code for the same amino acids are not all employed equally or at random in coding sequences; some are frequently selected over others.

This phenomenon of preferential or non-random use of synonymous codons is termed as codon usage bias (CUB) (Ma *et al.*, 2015; Liu, 2020), and it is a common occurrence in bacteria, plants, and mammals. Codon usage bias vary among genes within an organism and across different species, families, and taxonomic groups. Its evolution in various organisms is shaped by factors such as natural selection, genetic drift, and mutational pressure. Several key determinants influence the codon bias, including tRNA availability and interactions, recombination rates, mRNA secondary structure, codon position and context, GC content, gene expression levels, gene length, and overall genome composition (Parvathy *et al.*, 2022). The selection-mutation theory propounds that a finite population's equilibrium

between forces of selection and mutation produces a bias in the synonymous codon usage, with higher intragenomic bias signifying stronger selection for translation efficiency (Sharp *et al.*, 1986; Bulmer, 1991). On the other hand, the codon usage bias is said to result from distinct mutational pattern of synonymous codons, according to mutational or neutral theory. Thus, some alternative codons may experience a greater number of mutations than others, resulting in variations in codon utilization (Sharp *et al.*, 1993, Hershberg *et al.*, 2008). RNA editing sites are specific locations within RNA molecules where the original nucleotides are altered after transcription (the process of copying DNA into RNA).

Clarias batrachus (commonly known as Magur) and *Heteropneustes fossilis* (Singhi) are two economically and ecologically significant freshwater catfish species widely distributed across Asia. Their commercial value, combined with distinct adaptive traits, makes them valuable models for studying codon usage bias (CUB) in mitochondrial genes. Previous comparative studies on mitochondrial-encoded proteins in *C. batrachus*, particularly in relation to *C. gariepinus*, have revealed notable differences in protein structure, stability, and intrinsic disorder, suggesting possible molecular mechanisms underlying enhanced adaptability and competitive survival (Chand *et al.*, 2021). Similarly, *H. fossilis*, as a species adapted to diverse freshwater habitats, offers an opportunity to explore how evolutionary pressures and environmental conditions shape codon preference. Codon usage patterns are often influenced by factors such as mutational bias, natural selection, and translational efficiency, which collectively impact gene expression and organismal fitness. In this context, a comparative analysis of mitochondrial protein-coding genes in *C. batrachus* and *H. fossilis* can provide important insights into codon usage bias, focusing on identifying similarities and differences in the gene sequences, codon usage patterns, and phylogenetic relationships.

Materials and Methods

Sequence collection: The mitochondrial gene's coding sequences of two freshwater fish species (*Heteropneustes fossilis* and *Clarias batrachus*) were obtained from the GenBank database (<http://www.ncbi.nlm.nih.gov>) using accession numbers, i.e., NC_015827.1 for *Heteropneustes fossilis* and NC_023923.1 for *Clarias batrachus*. The CUB study of the mitochondrial genes of these two species was then analyzed by studying the following parameters.

CUB metrics

Compositional features of nucleotides: To analyse the compositional characteristics of mitochondrial genes in *Clarias batrachus* and *Heteropneustes fossilis*, the following parameters were examined: overall nucleotide composition (A%, T%, C%, and G%) and nucleotide distribution at the third codon position (T3%, A3%, C3%, and G3%); the proportion of guanine and cytosine in synonymous codons across each sequence, along with GC content at the first (GC1%), second (GC2%), and third

(GC3%) codon positions. Additionally, GC and AT skew values were calculated to assess the base complementary properties within each gene. All computational analyses were conducted using a PERL script developed by the corresponding author.

Effective Number of Codons (ENC): The CUB of a gene is commonly determined using the measure ENC, which ranges between 20 and 61. When the ENC value of a cds is 20, it reveals the highest codon bias implying that just one member out of a synonymous codon family is utilized to encode that specific amino acid in the polypeptide chain. In certain cases, synonymous codons for a given amino acid may be utilized uniformly with minimal or no bias in the coding sequence, resulting in an Effective Number of Codons (ENC) value of 61, indicating the absence of codon usage bias. However, a substantial degree of CUB is indicated by an ENC value less than or equal to 35 (Wright, 1990; Fuglsang, 2004).

The F_a value for every amino acid is used to determine the ENC. F_a was calculated by the formula:

$$F_a = \left(n_a \sum_{i=1}^k p_i^2 - 1 \right) / (n_a - 1)$$

Where, k denotes the number of synonymous codons for the a^{th} amino acid, n_a denotes the observed codon value for the amino acid, and p_i is the frequency of the i^{th} codon. The following formula may be used to evaluate the mean of the F_a values for each r -fold redundancy class of synonymous codons (folds 2, 4, and 6 for amino acids in NCBI's translation, Table 2):

$$\bar{F}_r = \frac{1}{nRC} \sum_{a \in RC} F_a$$

Where, the total amount of amino acids is indicated by nRC . ENC can be computed as follows:

$$ENC = \left(\frac{12}{\bar{F}_k} \right) + \left(\frac{6}{\bar{F}_k} \right) + \left(\frac{2}{\bar{F}_k} \right)$$

In k -fold degenerate amino acids, the average of the F_k values ($k=2, 4$ and 6) is represented by the symbol \bar{F}_k .

Relative Synonymous Codon Usage (RSCU): RSCU is a metric for identifying the codons that appear more often than expected within a gene. It is calculated as the ratio of observed frequency of a codon to its expected frequency under random and equal usages. RSCU value greater than 1.0 signifies that a codon is favoured and used more often than expected. Conversely, RSCU value less than 1.0 suggests that a codon is used less frequently than expected. Within the coding sequences, the codons with RSCU values exceeding 1.6 are considered over-represented codons, while those with values below 0.6 are considered as under-represented ones (Sharp et al., 1986). The RSCU value of a codon was calculated by the formula:

$$RSCU = \frac{X_{ij}}{\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}} s$$

In this case, X_{ij} is the frequency of the j^{th} codon for the i^{th} amino acid, and n_i is the number of codons for the i^{th} amino acid. A random value of 0.5 is assigned to any X_{ij} that has a value of zero.

Correspondence analysis (COA): This computational tool analyses large-scale patterns of variation to understand how often the synonymous codons are used within the mitochondrial genes. It categorizes codons into two groups based on the trends. In the context of the mitochondrial gene, each gene can be represented as a vector with 60 dimensions. Each dimension corresponds to the RSCU value of a single synonymous codon (out of a total of 60 possible codons). A concept called relative inertia can be applied to these 60-dimensional vectors. It helps identifying the underlying forces shaping the codon usage patterns based on the location of genes within the mitochondrial genome (Shields et al., 1987, Singh et al., 2016).

Pr2 (Parity plot) bias analysis: Scatter plot depicts the influence of two evolutionary factors on codon usage bias. The vertical axis (ordinate) represents the At_{bias} , i.e. $[A3/ (A3+T3)]$ whereas the horizontal axis (abscissa) represents the GC_{bias} , $[G3/ (G3+C3)]$. The center of the plot corresponds to a value of 0.5 on both the axes. This position signifies a scenario where the abundance of $A=T$ and $G=C$. In this balanced state, the two evolutionary factors, presumably favouring AT or GC composition, have minimal impact (Sueoka, 1995).

Neutrality Plot: This plot allows us to assess the relative influence of two evolutionary forces on codon usage bias (CUB) (Zhang et al., 2013). The vertical axis (y-axis) represents GC12 value, i.e., mean GC contents at 1st and 2nd codon positions. The horizontal axis (x-axis) represents the GC3 content of a gene. The slope (regression coefficient) of a straight line in this plot provides useful clues about the dominant force shaping the CUB. A slope closer to 1 suggests that mutational pressure has a stronger influence on CUB of gene. Conversely, a slope near 0 indicates that natural selection, rather than mutational pressure, plays a more prominent role in shaping the CUB (Sharp et al., 1986; Sueoka 1995).

Protein properties: The aromaticity, hydrophilicity and GRAVY (grand average of hydropathicity) of the protein encoded by the coding sequence was determined. A positive GRAVY score represents a hydrophobic protein, whereas a hydrophilic protein is represented by a negative GRAVY score (Kyte et al., 1982). The aromaticity of a protein refers to how frequently the complex structured amino acids namely Tyr, Trp and Phe occur in the protein (Lobry and Gautier, 1994). Hydrophilicity of a protein defines the degree of hydrophilic (water-loving) character of the amino acids of that protein.

Study of phylogenetics: The evolutionary relationships between the mitochondrial genes of *C. batrachus* and *H. fossilis* were investigated using phylogenetic tree constructed with MEGA 11 software (Munjal et al., 2019).

RNA editing sites: RNA editing refers to a post-transcriptional process that modifies the nucleotide sequence of RNA molecules after they are transcribed. Two primary forms of RNA editing have been characterized: the enzymatic conversion of adenosine to inosine (A-to-I) and cytidine to uridine (C-to-U) (Karagianni, 2024).

Data analysis: Correlations between various parameters and codon usage bias (CUB) in the genes were analysed by a statistical software, specifically SPSS 21.0 for Windows. These parameters included base composition, effective number of codons, nucleotide skew values, and protein properties. To understand the overall codon usage pattern, correspondence analysis (COA) was performed. Additionally, PAST software was used to investigate variations in the compositional distribution of synonymous codons.

Results and Discussion

Nucleotide compositional properties: According to recent CUB research, the total nucleotide composition of genes affects the utilization of synonymous codons (Zhao et al., 2024). This study revealed the base composition and total nucleotide abundance at the third codon position of the genes in both the species (Fig. 1). The proportions of bases T (27% in *C. batrachus* and 28% in *H. fossilis*) and C (27% in *C. batrachus* and 26% in *H. fossilis*) were slightly different. The base compositions of A (~29%) and G (~15%) were observed in similar proportions in both *C. batrachus* and *H. fossilis*. The mitochondrial genes of *H. fossilis* and *C. batrachus* were, in fact, AT-rich, as evidenced by their respective overall GC% contents of 41.90% and 42.68% and their overall AT% of 58.10% and 57.32% for both the mt-genomes. In terms of frequency, A (44.74% and 44.39%) was found to be the most common nucleotide amongst the two fish species, followed by C (28.96% and 27.54%), T (18.92% and 22.31%), and G (7.38% and 5.76%). For *C. batrachus* and *H. fossilis*, the respective total GC3 concentrations were 36.35% and 33.32%, while the corresponding AT3 levels were 63.65% and 66.68%. This suggests that the third codon location had a high AT content in the mt-genes of both fish species. In the mitochondrial genes of two fish species, the pattern of GC concentration across all codon positions was GC1>GC2>GC3 (Fig. 2). Table 1 shows the findings of the link between ENC and each nucleotide base. In *Clarias batrachus*, the correlation between the ENC and T%, C%, A3%, C3%, GC% and GC3% showed a positive correlation whereas ENC with A%, G%, T3% and G3% showed negative correlation. In *Heteropneustes fossilis*, the correlation between ENC and C%, C3%, GC% and GC3% showed positive correlation whereas ENC with A%, T%, G%, A3%, T3% and G3% showed negative correlation. It can be inferred that the base composition affects the CUB of genes based on the correlation results.

Barbhuiya et al. (2021) investigated the base composition of mitochondrial CO genes in different amphibian groups (tailed amphibians, legless amphibians, and frogs). Interestingly, the order of base frequency varied between the groups. In tailed and

legless amphibians, adenine (A) was most frequent, followed by thymine (T), cytosine (C), and guanine (G). However, in frogs, the order switched, with adenine being the most frequent, followed by cytosine, thymine and guanine. This difference suggests variation in AT richness across the groups. The researchers also observed that the total amount of adenine and thymine (AT) across the CO genes was consistently higher than guanine and cytosine (GC) in all groups, indicating an overall bias towards AT nucleotides. Notably, frogs had the highest GC content, while tailed amphibians had the lowest. Based on these findings, the study concluded that tailed amphibians possessed the highest AT percentage, while frogs had the lowest. Legless amphibians fell in between in terms of GC and AT contents.

ENC study: ENC values were calculated for the mitochondrial genes of *C. batrachus* and *H. fossilis* to study the CUB. In *C. batrachus*, the ENC values ranged between 35-61 with a mean of 43.46. In *H. fossilis*, the ENC values varied between 36-61 with a mean value of 44.08. As the ENC values were greater than 35, it indicated that the CUB of mitochondrial genes in these two species was low (Butt et al., 2014). Deb et al. (2020) discovered that the mean ENC value was greater than 35 in various genomes of hepadnaviruses, which corroborates with the findings of this study.

RSCU (Relative Synonymous Codon Usage): In this study, 12 over-represented codons were discovered in *C. batrachus* with RSCU value >1.6. On the other hand, 19 codons were under-represented with RSCU value < 0.6 whereas 26 codons were found to be the preferred ones with RSCU value >1. However, 11 codons were found in *H. fossilis* that were over-represented with RSCU value >1.6; 17 under-represented codons with RSCU value < 0.6, and 23 preferred codons with RSCU value >1 (Table 2, 3; Fig. 3).

Deb et al. (2021) analyzed the RSCU values for each member of Family Anelloviridae and discovered that 27 out of 59 codons were used more frequently. They had noted a preference for codons ending in A and C. Specifically, the AGA codon was over-represented, while TTG, TCG, CGT, CGG, GCG, GAT and ACG codons were consistently under-represented across all the studied genomes (Deb et al., 2021).

Correspondence analysis: The codons encoding sense strands in mitochondrial genes of *C. batrachus* and *H. fossilis* were used for the correspondence analysis (COA) (Fig. 4). Significant contributions to the total variance are made by both the axes. In *C. batrachus*, axis 1 occupied 57.06% of the total variance whereas axis 2 occupied 11.83%. In *H. fossilis*, axis 1 occupied 44.44% of the total variance, while axis 2 occupied 16.37%. In this study, GC-ending codons were represented by green dots, while AT-ending codons by red dots (Fig. 4). The codon usage bias (CUB) in the mitochondrial genes of both the fish species appeared to be influenced by mutational pressure, as suggested by the close clustering of bases along the axes in the figures. Deb et al. (2018) found that mitochondrial ND gene codons in amphibians clustered closely along both the axes, indicating mutational pressure as a key driving force of codon usage bias (CUB), aligning with the findings of this study in two fish species.

Table 1: Correlation between ENC and base contents of genes in *Clarias batrachus* and *Heteropneustes fossilis*

		A%	T%	G%	C%	A3%	T3%	G3%	C3%	GC%	GC3%
ENC	<i>Clarias batrachus</i>	-0.199	0.070	-0.076	0.255	0.035	-0.119	-0.077	0.155	0.341	0.160
	<i>Heteropneustes fossilis</i>	-0.017	-0.086	-0.173	0.272	-0.002	-0.190	-0.107	0.271	0.226	0.438

**Significant at 0.01 level (2-tailed); *Significant at 0.05 level (2-tailed)

Table 2: Over-represented and under-represented codons of *Clarias batrachus* and *Heteropneustes fossilis*

<i>Clarias batrachus</i>		<i>Heteropneustes fossilis</i>	
Over-represented codons	Under-represented codons	Over-represented codons	Under-represented codons
TCA	TCG	TCA	TCG
CTA	AGT	CTA	AGT
CCA	TTG	CCA	TTG
CAA	CTG	CAA	CTG
CGA	TGT	CGA	CCG
TGA	CCG	TGA	CAG
ACA	CCT	ACA	CGG
AAC	CAT	AAA	CGT
GTA	CAG	GTA	TGG
GCA	CGG	GAA	ACG
GAA	TGG	GGA	ACT
GGA	ACG	-	AAG
-	ACT	-	GTG
-	AAG	-	GCG
-	GTG	-	GAG
-	GTT	-	GGG
-	GCG	-	ATG
-	GAT	-	-

Chakraborty *et al.* (2020) investigated the uneven distribution of nucleotides (guanine-cytosine, *i.e.*, GC and adenine-thymine *i.e.*, AT) in obesity and housekeeping genes using parity plot analysis. Their findings suggest that two factors, mutational pressure and natural selection, might play a role in shaping the base composition (CUB) of mitochondrial genes in two species (Chakraborty *et al.*, 2020).

Interrelationships among the nucleotide compositions:

Mutational pressure and natural selection are the two main evolutionary forces responsible for codon usage bias (Mazumder *et al.*, 2018 a,b). To determine the main factors affecting the CUB of genes, a correlation study of the nucleotide compositional features was performed using Karl Pearson's method (Chen 2013). A highly significant correlation ($p < 0.01$ or 0.05) between total nucleotide content and nucleotide composition at the third codon position (Table 4) was identified. This finding suggests that mutational pressure may play a key role in shaping codon usage bias (CUB) in the mitochondrial genes of two fish species (Zhao *et al.*, 2007; Zhang *et al.*, 2013).

Chakraborty *et al.* (2020) found a strong positive correlation between nucleotide content and third codon position

in obesity-related genes, suggesting mutational pressure as a key driver of codon usage bias. However, a significant G/C and C3/T3 correlation indicated the role for natural selection. In housekeeping genes, the mixed positive and negative correlations suggested the combined influence of mutational pressure and natural selection.

Neutrality plot: It is inspected to quantify the impact of evolutionary pressures on the mt-genes in addition to the correlation between GC3 and GC12. A regression equation was generated by plotting GC3 and GC12 on the x and y-axis, respectively, for the mt genes of both species (Fig. 6). When the regression coefficient (RC) value is less than 0.5, natural selection was generally found to be more significant than mutational pressure; conversely, when the RC value was more than 0.5, mutational pressure is deemed to be more significant. Our research revealed RC values of 0.599 for *H. fossilis* and 0.221 for *C. batrachus*, suggesting that natural selection was more influential than mutational pressure in defining the CUB of mt genes in *C. batrachus*. Whereas in *H. fossilis*, the RC value indicates that both mutation pressure and natural selection shape CUB in the mt-genes, with mutation having a slightly stronger influence. Abdoli *et al.* (2022) found an RC value < 0.5 for GC12 on

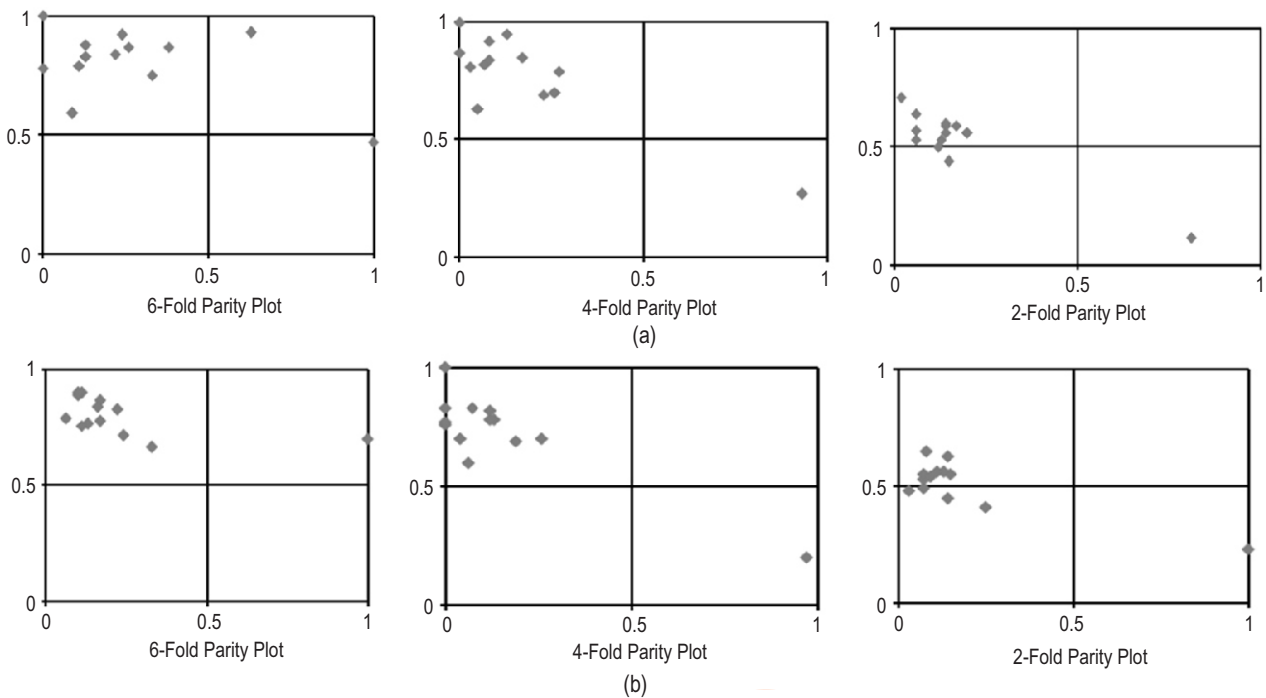


Fig. 5: Parity bias plots of 2, 4 and 6-fold degenerate codons in (A) *Clarias batrachus* and (B) *Heteropneustes fossilis*.

GC3 in all 13 mitochondrial genes of five silkworm species, indicating that natural selection, rather than mutational pressure, played a dominant role in shaping codon usage bias (CUB) (Abdoli et al., 2022).

Nucleotide skewness: In this experiment, *H. fossilis* had an average AT skew of 0.019 and an average GC skew of -0.289, whereas *C. batrachus* had an average AT skew value of 0.038 and an average GC skew value of -0.285. Both A and C nucleotides were used more frequently than T and G nucleotides in the coding regions of the mitochondrial genes of two fish species (Wei et al., 2014a). Nucleotide skewness was found to be a significant factor in determining the CUB of genes or genomes in previous studies. The connection between ENC and nucleotide skews in *C. batrachus* was evaluated. The results showed that ENC had a positive correlation with keto skew (0.117), but a negative correlation with AT, PU, amino, GC and pyrimidine (PY) skews (-0.142, -0.037, -0.597*, -0.115 and -0.102). Conversely, it was noted that the ENC of *H. fossilis* exhibited a negative correlation with GC skew (-0.209), pyrimidine skew (-0.197) and amino skew (-0.427), but a positive correlation with AT skew (0.020), purine skew (0.099) and keto skew (0.217). It can be inferred from these results that nucleotide skewness may have affected the CUB of mt-genes. Recent research has shed light on the uneven distribution of nucleotides in various viruses. Deb et al. (2020) observed a preference for cytosine (C) over thymine (T) and guanine (G) over adenine (A) in hepadnaviruses. Their analysis revealed a negative correlation between AT content and several other measures of nucleotide bias, except for keto skew

Table 3: Preferred codons of *Clarias batrachus* and *Heteropneustes fossilis*

Preferred codons	
<i>Clarias batrachus</i>	<i>Heteropneustes fossilis</i>
TCA	TCA
TCC	TCC
TCT	AGC
TTC	TTC
TTA	TTA
CTA	CTA
TAC	CCA
TGC	CAC
CCA	CAA
CCC	CGA
CAC	TGA
CAA	ATA
CGA	ATT
TGA	ACA
ATA	ACC
ATT	AAC
ACA	AAA
ACC	GTA
AAC	GCA
AAC	GCC
GTA	GAC
GCA	GAA
GCC	GGA
GAC	-
GAA	-
GGA	-

Table 4: Correlation study in *Clarias batrachus* and *Heteropneustes fossilis* between overall nucleotide contents and nucleotide contents in the third position of codons

	<i>Clarias batrachus</i>					<i>Heteropneustes fossilis</i>				
	A3%	T3%	G3%	C3%	GC3	A3%	T3%	G3%	C3%	Gc3
A	0.922**	-0.843**	-0.840**	0.604*	-0.645*	0.858**	-0.917**	-0.822**	0.800**	0.107
T	-0.927**	0.929**	0.961**	-0.809**	0.480	-0.901**	0.953**	0.902**	-0.867**	-0.087
G	-0.924**	0.936**	0.966**	-0.825**	0.457	-0.914**	0.969**	0.963**	-0.926**	-0.096
C	0.781**	-0.884**	-0.948**	0.928**	-0.192	0.855**	-0.898**	-0.946**	0.899**	0.069
GC	-0.601*	0.421	0.362	-0.070	0.689**	-0.423	0.475	0.326	-0.342	-0.106

Table 5: Correlation analysis of the protein properties with ENC

	ENC	Gravy	Hydrophilicity	Aromaticity
<i>Clarias batrachus</i>		0.251	-0.162	-0.043
<i>Heteropneustes fossilis</i>		-0.090	-0.121	0.158

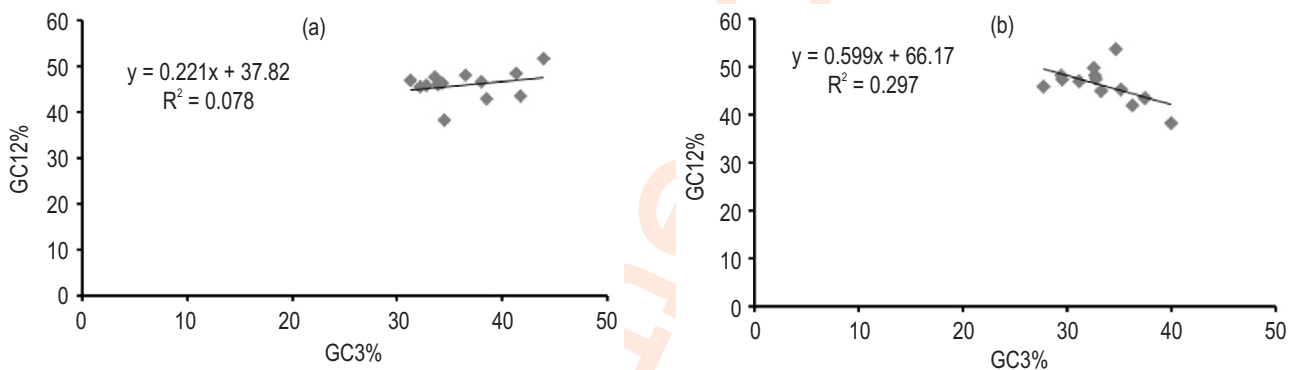


Fig. 6: Neutrality plots of (A) *Clarias batrachus* and (B) *Heteropneustes fossilis*.

which showed a positive correlation (Deb *et al.*, 2020). Additionally, Chakraborty *et al.* (2019) identified a significant link between codon usage and nucleotide imbalances in Nipah virus genes. These findings suggest that different types of viruses exhibit distinct patterns in their nucleotide composition (Chakraborty *et al.*, 2019).

Protein properties: Pearson's correlation analysis was applied to investigate the relationship between CUB and GRAVY, as well as the aromaticity and hydrophilicity of mt-genes of *C. batrachus* and *H. fossilis*. It was evident from this study that in *C. batrachus*, the protein GRAVY showed a positive correlation with ENC and hydrophilicity and aromaticity showed a negative correlation with ENC whereas protein aromaticity showed a positive correlation with the ENC, and GRAVY as well as hydrophilicity showed a negative correlation with ENC in *H. fossilis* (Table 5). Uddin *et al.* (2020) in their studies showed that the proteins generated by the ATP6 and ATP8 genes in fish, birds, and mammals have higher concentrations of amino acids such as cysteine, threonine, and

alanine (Uddin *et al.*, 2020).

Phylogenetic study: The results of the phylogenetic clustering are mentioned in Fig. 7. For both the fish species, Neighbor-Joining method was used to create a family tree based on the genetic differences of the mitochondrial genes between them (Saitou *et al.*, 1987). In *C. batrachus*, the length of the branches on the tree reflects the number of mutations that likely occurred in the DNA at those points. The study includes 1830 segments of DNA, including both coding and non-coding regions. Any unclear or uncertain sections of DNA were excluded from the analysis. For *Heteropneustes fossilis*, the lengths of the branches on the tree reflect the degree of evolutionary changes between the coding sequences. These changes were measured as the number of substitutions in the building blocks (nucleotides) of the genetic material. The analysis considered all parts of the genetic code (codons) as well as non-coding regions. Any unclear section in the sequences was excluded from the comparison. The final

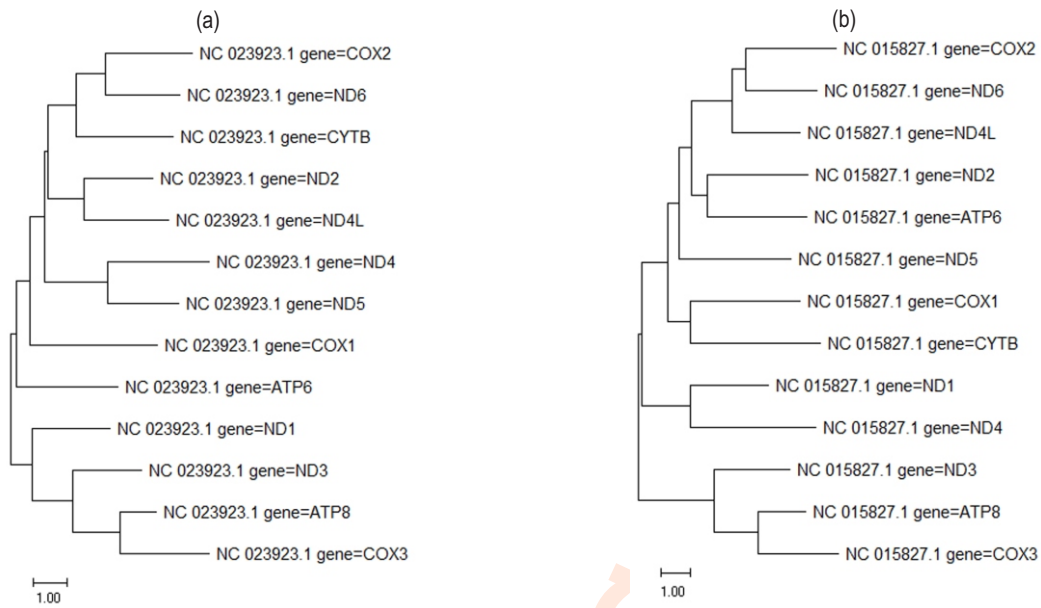


Fig. 7: Comparative study of phylogenetic tree of (A) *Clarias batrachus* and (B) *Heteropneustes fossilis*.

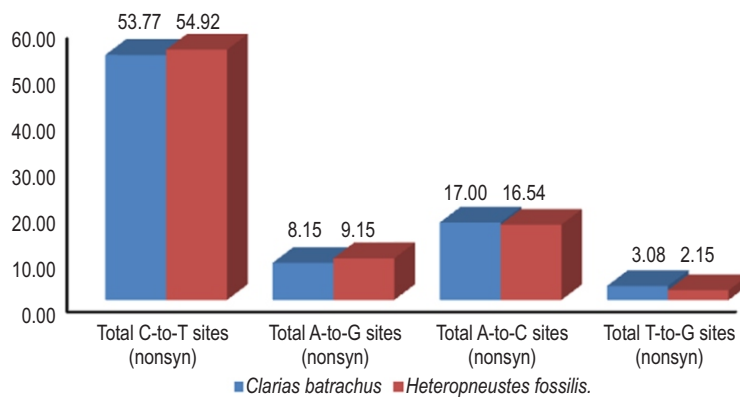


Fig. 8: Comparative study of RNA editing sites of *C. batrachus* and *H. fossilis*.

dataset included 1824 positions. The results of the phylogenetic clustering, presented in Fig. 7, ideally showed that the mt-genes of *C. batrachus* and *H. fossilis* were closely related.

RNA editing sites: RNA editing sites play a vital role in enhancing both transcriptome and proteome variability, enabling precise regulation of gene expression and modification of protein functionality. Analysis of RNA editing sites in *C. batrachus* (53.77%) and *H. fossilis* (54.92%), revealed high rates of C to T substitutions (non-synonymous) (Fig. 8). This suggests a rapidly evolving lineage, possibly due to adaptation to environmental changes, selective pressure, or positive selection for specific mutations. Our analysis revealed distinct codon usage patterns in the mitochondrial genes of *C. batrachus* and *H. fossilis*, with notable over- and under-

represented synonymous codons. The observed AT-rich bias at third codon position and the influence of natural selection and mutation suggest adaptive evolution in response to their aquatic environments. These findings enhance our understanding of mitochondrial genome evolution in catfish species.

Acknowledgment

All the authors duly acknowledge the Department of Biotechnology, Assam University, Silchar for providing necessary laboratory facilities.

Authors' contribution: J.C. Das: Original draft, Analysis of data, Investigation, Interpretation, Writing-Revision and Editing,

Validation; **V. Dey**: Original draft, Data curation, Analysis of data, Investigation, Interpretation, Writing and Editing, Validation; **S. Chakraborty**: Conceptualization, Project administration, Formal analysis, Software, Visualization, Supervision, Methodology, Writing- Revision and Editing.

Funding: The work was unfunded.

Research content: This work is based on codon usage bias (CUB) analysis and the RNA editing sites on the mitochondrial genes.

Ethical approval: Not applicable. This study is based on DNA sequence analysis and the data were retrieved from publicly available databases from NCBI.

Conflict of interest: The authors have declared that no conflict of interest exists for this research work.

Data availability: The datasets generated and/or analysed during the current study are available in the NCBI repository (<https://www.ncbi.nlm.nih.gov/>).

Consent to publish: All the authors of the manuscript have confirmed and agreed to the terms and conditions of publication as outlined by the Journal.

References

- Abdoli, R., T.H. Mazumder, S. Nematollahian, R.S. Zanjani, R.A. Mesbah and A. Uddin: Gaining insights into the compositional constraints and molecular phylogeny of five silkworms mitochondrial genome. *Int. J. Biol. Macromole.*, **206**, 543-552 (2022).
- Barbhuiya, P.A., A. Uddin and S. Chakraborty: Codon usage pattern and evolutionary forces of mitochondrial ND genes among orders of class Amphibia. *J. Cellu. Physiol.*, **236**, 2850-2868 (2021).
- Barbhuiya, P.A., A. Uddin and S. Chakraborty: Understanding the codon usage patterns of mitochondrial CO genes among Amphibians. *Gene*, **777**, 145462 (2021).
- Bulmer, M.: The selection-mutation-drift theory of synonymous codon usage. *Genetics*, **129**, 897-907 (1991).
- Butt, A.M., I. Nasrullah and Y. Tong: Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. *PLoS One*, **9**, e90905 (2014).
- Chakraborty, S., P.A. Barbhuiya, S. Paul, A. Uddin, Y. Choudhury, Y. Ahn and Y.S. Cho: Codon usage trend in genes associated with obesity. *Biotechnol. Lett.*, **42**, 1865-1875 (2020).
- Chakraborty, S., B. Deb, P.A. Barbhuiya and A. Uddin: Analysis of codon usage patterns and influencing factors in Nipah virus. *Virus Res.*, **263**, 129-138 (2019).
- Chand, G.B., S. Kumar and G.K. Azad: Molecular assessment of proteins encoded by the mitochondrial genome of *Clarias batrachus* and *Clarias gariepinus*. *Biochemi. Biophys. Rep.*, **26**, 100985 (2021).
- Chen, Y.: 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. Res. Int.*, **2013**, 406342 (2013).
- Collen, B., F. Whitton, E.E. Dyer, J.E. Baillie, N. Cumberlidge, W.R. Darwall, C. Pollock, N.I. Richman, A.M. Soulsby and M. Böhm: Global patterns of freshwater species diversity, threat and endemism. *Glob. Ecol. Biogeogr.*, **23**, 40-51 (2014).
- Deb, B., A. Uddin and S. Chakraborty: Codon usage pattern and its influencing factors in different genomes of hepadnaviruses. *Arch. Virol.*, **165**, 557-570 (2020).
- Deb, B., A. Uddin and S. Chakraborty: Composition, codon usage pattern, protein properties, and influencing factors in the genomes of members of the family Anelloviridae. *Arch. Virol.*, **166**, 461-474 (2021).
- Deb, B., A. Uddin, G.A. Mazumder and S. Chakraborty: Analysis of codon usage pattern of mitochondrial protein-coding genes in different hookworms. *Mole. Bioche. Parasitol.*, **219**, 24-32 (2018).
- Diogo, R., M. Chardon and P. Vandewalle: On the osteology and myology of the cephalic region and pectoral girdle of *Heteropneustes fossilis* (Siluriformes: Heteropneustidae), with comments on the phylogenetic relationships between *Heteropneustes* and the clariid catfishes. *Anim. Biol.*, **53**, 379-396 (2003).
- Fuglsang, A.: The 'effective number of codons' revisited. *Bioche. Biophys. Res. Commun.*, **317**, 957-964 (2004).
- Hasan, M.R., M.Y. Hossain, Z. Mawa and M.A.R. Hossain: Reproductive biology of *Heteropneustes fossilis* in a wetland ecosystem (Gajner Beel, Bangladesh) in relation to eco-climatic factors: Suggesting a sustainable policy for aquaculture, management and conservation. *Saudi J. Biol. Sci.*, **29**, 1160-1174 (2022).
- Hershberg, R. and D.A. Petrov: Selection on codon bias. *Ann. Rev. Genet.*, **42**, 287-299 (2008).
- Karagianni, K., A. Bibi, A. Madé, S. Acharya, M. Parkkonen, T. Barbalata, P. K. Srivastava, D. Gonzalo-Calvo, C. Emanuelli, F. Martelli, Y. Devaux, D. Dafou and A. Y. Nossent: Recommendations for detection, validation, and evaluation of RNA editing events in cardiovascular and neurological/neurodegenerative diseases. *Molecular Therapy Nucleic Acids*, **35**, 102085 (2024).
- Kumar, A., K. Jitender and K. Hemendra: Length-weight relationship and relative condition factor of *Clarias batrachus* (Linnaeus, 1758) from Gaurmati Fish Farm, Kawardha, Chhattisgarh, India. *Int. J. Curr. Microbiol. Appl. Sci.*, **6**, 1425-1431 (2017).
- Kushwaha, B., R. Kumar, S. Agarwal, M. Pandey, N.S. Nagpure, M. Singh, S. Srivastava, C.G. Joshi, P. Das, L. Sahoo, P. Jayasankar, P.K. Meher, T.M. Shah, A.B. Patel, N. Patel, P. Koringa, S.P. Das, S. Patnaik, A. Bit, Sarika, M.A. Iquebal, D. Kumar and J.K. Jena: Assembly and variation analyses of *Clarias batrachus* mitogenome retrieved from WGS data and its phylogenetic relationship with other catfishes. *Meta Gene*, **5**, 105-114 (2015).
- Li, N., L. Bao, T. Zhou, Z. Yuan, S. Liu, R. Dunham, Y. Li, K. Wang, X. Xu, Y. Jin, Q. Zeng, S. Gao, Q. Fu, Y. Liu, Y. Yang, Q. Li, A. Meyer, D. Gao and Z. Liu: Genome sequence of walking catfish (*Clarias batrachus*) provides insights into terrestrial adaptation. *BMC Genomics*, **19**, 952 (2018).
- Liu, Y.: A code within the genetic code: codon usage regulates co-translational protein folding. *Cell Commun. Signal.*, **18**, 145 (2020).
- Ma, Q., C. Li, J. Wang, Y. Wang and Z. Ding: Analysis of synonymous codon usage in FAD7 genes from different plant species. *Genet. Mol. Res.*, **14**, 1414-1422 (2015).
- Mazumder, G.A., A. Uddin and S. Chakraborty: Codon usage pattern of complex III gene of respiratory chain among Platyhelminths. *Gene. Evol.*, **57**, 128-137 (2018).
- Mazumder, G.A., A. Uddin and S. Chakraborty: Preference of A/T ending codons in mitochondrial ATP6 gene under phylum Platyhelminthes: codon usage of ATP6 gene in Platyhelminthes. *Mole. Bioche. Parasitol.*, **225**, 15-26 (2018).
- Montaña-Lozano, P., M. Moreno-Carmona, M. Ochoa-Capera, N.S.

- Medina, J.L. Boore and C.F. Prada: Comparative genomic analysis of vertebrate mitochondrial reveals a differential of rearrangements rate between taxonomic class. *Scie. Rep.*, **12**, 5479 (2022).
- Munjaj, G., M. Hanmandlu and S. Srivastava: Phylogenetics algorithms and applications. *Ambi. Communicat. Compu. Syst.*, **904**, pp. 187–194 (2019). doi: 10.1007/978-981-13-5934-7_17.
- Parvathy, S.T., V. Udayasuriyan and V. Bhadana: Codon usage bias. *Mol. Biol. Rep.*, **49**, 539-565 (2022).
- Puvaneswari, S., K. Marimuthu, R. Karuppasamy and M.A. Haniffa: Early embryonic and larval development of Indian catfish, *Heteropneustes fossilis*. *EurAsian J. BioSci.*, **3**, 84-96 (2009).
- Ratmuangkhwang, S., P. Musikasinthorn and Y. Kumazawa: Molecular phylogeny and biogeography of air sac catfishes of the *Heteropneustes fossilis* species complex (Siluriformes: Heteropneustidae). *Mole. Phylogene. Evolut.*, **79**, 82-91 (2014).
- Saitou, N. and M. Nei: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406-425 (1987).
- Sharp, P.M. and W.H. Li: An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mole. Evolu.*, **24**, 28-38 (1986).
- Sharp, P.M. and W.H. Li: An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.*, **24**, 28-38 (1986).
- Sharp, P.M., M. Stenico, J.F. Peden and A.T. Lloyd: Codon usage: mutational bias, translational selection, or both? *Bioche. Soci. Transac.*, **21**, 835-841 (1993).
- Shields, D.C. and P.M. Sharp: Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. *Nucl. Acids Res.*, **15**, 8023-8040 (1987).
- Singh, N.K., A. Tyagi, R. Kaur, R. Verma and P.K. Gupta: Characterization of codon usage pattern and influencing factors in Japanese encephalitis virus. *Virus Res.*, **221**, 58-65 (2016).
- Sueoka, N.: Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. *J. Mol. Evol.*, **40**, 318-325 (1995).
- Tamura, K., G. Stecher and S. Kumar: MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.*, **38**, 3022-3027 (2021).
- Uddin, A., T.H. Mazumder, P.A. Barbhuiya and S. Chakraborty: Similarities and dissimilarities of codon usage in mitochondrial ATP genes among fishes, aves, and mammals. *IUBMB Life*, **72**, 899-914 (2020).
- Uddin, A., T.H. Mazumder, M.N. Choudhury and S. Chakraborty: Codon bias and gene expression of mitochondrial ND2 gene in chordates. *Bioinformatics*, **11**, 407-412 (2015).
- Wei, L., J. He, X. Jia, Q. Qi, Z. Liang, H. Zheng, Y. Ping, S. Liu and J. Sun: Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. *BMC Evolution. Biol.*, **14**, 1-12 (2014).
- Wright, F.: The 'effective number of codons' used in a gene. *Gene*, **87**, 23-29 (1990).
- Zhang, Z., W. Dai and D. Dai: Synonymous codon usage in TTSuV2: analysis and comparison with TTSuV1. *PLOS ONE*, **8**, e81469 (2013).
- Zhang, Z., W. Dai, Y. Wang, C. Lu and H. Fan: Analysis of synonymous codon usage patterns in torque teno sus virus 1 (TTSuV1). *Arch. Virol.*, **158**, 145-154 (2013).
- Zhao, S., Q. Zhang, Z. Chen, Y. Zhao and J. Zhong: The factors shaping synonymous codon usage in the genome of *Burkholderia mallei*. *J. Gene. Genom.*, **34** 362-372 (2007).
- Zhao, Y. and S. Zhang: Comparative analysis of codon usage bias in six *Eimeria* genomes. *Int. J. Mole. Sci.*, **25**, 8398 (2024).