IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Translational Efficiency And Mutational Drift Of Cyb Gene In Balitoridae Fish Family

[1]Joydeep Chandra Das, [1]Supriyo Chakraborty*

Affiliation: [1]Department of Biotechnology, Assam University, Silchar-788011, Assam, India

Abstract

Study of molecular biology of fishes sheds light on the mechanisms guiding evolution and forming the diversity of fishes on Earth. It contributes to the understanding of the genetic foundations of evolutionary change, the creation of new species, and the adaptation to diverse environments. Studying mitochondrial genes of an organism using codon usage bias (CUB) is important because it helps reveal the patterns of evolution in the context of molecular biology. In this study on the MT-CYB gene of Balitoridae family, we described the codon usage patterns and the variables that influenced it. According to correspondence analysis (COA), axis-1 reflected 43.05% of the codon usage variation, whereas axis-2 occupied 16.70% of the total variation. GC content in three codon positions of CYB gene was found in the order GC1>GC3>GC2. With a ratio of 0.0768/0.9232=0.08, the neutrality plot demonstrated that both the evolutionary forces (mutation and natural selection) had an effect on the MT-CYB gene. Natural selection was shown to be the most important factor, while compositional constraints and mutation pressure were also significant in influencing the pattern of codon usage. Thirteen codons were found to be over-represented, whereas twenty-four codons were found to be under-represented as revealed by the RSCU values of the codons in mitochondrial cytochrome B gene of Balitoridae. Analysis of codon usage patterns increased our comprehension of the mechanisms driving the codon distribution and the evolutionary significance of the MT-CYB gene.

Keywords: Codon usage bias; MT-CYB gene; Balitoridae family; natural selection; mutational pressure

1. Introduction

Balitoridae is a family of tiny freshwater fishes which include hill stream loaches and river loaches from East, South and Southeast Asia. They belong to the order Cypriniformes and there are around 202 species in the family. They are sometimes referred to as ‘lizard fish’ or ‘flossensaugers’ (in Germany). Many of the species are popular in aquaria, with the *Sewellia* genus being the most regularly marketed in the aquaria trade (Randall and Page 2015). They share several characteristics with the family Cobitidae, their sibling family of ‘loaches’, including several barbels surrounding the mouth (Kottelat 2012). The majority of species live in rapid, clean and well-oxygenated streams and are rheophilic. A few members of the family have modified ventral fins that they employ to cling to rocks in fast-moving streams or torrents (Paxton and Eschmeyer 1998).

Studies on codon usage bias in fish molecular biology is essential to comprehend the evolutionary processes influencing genetic diversity and adaptation. Understanding codon usage bias helps one better understand the ways that genetic drift, mutation pressure, and natural selection affect the evolution of genomes (Behura and Severson 2012). Analysing codon preferences in fishes can provide insight into how they have adapted to different ecological niches and environmental stresses, like temperature and oxygen availability. This approach improves our comprehension of the molecular processes that underlie fish evolution and their adaptation to shifting environmental conditions.

The genetic code is degenerate; therefore there exists several choices for codons of a particular amino acid. Preferences for synonymous codons differ among the species, and this occurrence of biased codon selection resulting from evolution is referred to as codon usage bias (CUB) (Bulmer 1991). All organisms experience distinct variations of this phenomenon, and sometimes the same organism's genome has multiple CUBs in different regions (Hershberg and Petrov 2008). According to earlier studies on codon usage; natural selection, mutation and genetic drift appear to have been the main influencers on codon usage bias (Bulmer 1991). There is a widespread understanding that the main factor underlying codon bias selection is effective protein synthesis by optimum codons that promote fitness (Stenico, Lloyd et al. 1994). GC contents, tRNA abundance, protein hydropathy, protein structure, gene expression level, gene length, recombination, environmental stress, etc. are some additional elements that also influence such bias. Codon usage research has been extensively undertaken to better understand genome evolution, heterologous gene expression and to design degenerate primers (Sharp and Matassi 1994, Zhou, Gu et al. 2005). Earlier studies examined the pattern of codon usage in a variety of species including complex eukaryotes and also prokaryotes like viruses, bacteria, etc. (Akashi 1997, Shackelton, Parrish et al. 2006). It is thought that the nucleotide composition has the most influence on codon usage bias (Shang, Liu et al. 2011). The literature has reported on the relation between synonymous codon usage and GC multiple times. The average G+C composition of complete genomes can vary significantly, which is a well-known phenomenon (Muto and Osawa 1987). Furthermore, some genes from various genomes may have different G+C compositions (Jermiin, Graur et al. 1994). Both synonymous and non-synonymous codon sites are equally susceptible to variations in G+C composition (Jukes and Bhushan 1986). In the previous studies, synonymous alterations were assumed to be neutral in terms of protein function and to have no impact. However, current research indicates that synonymous sites function to influence the expression and structure of proteins (Seligmann and Warthi 2017). Synonymous mutations are associated with cancer (Schutz, Pomerantz et al. 2013) and many other common disorders (Sauna and Kimchi-Sarfaty 2011).

There are differences between the standard genetic code and the mitochondrial genetic code (Table 1). The 13 protein-coding genes, 2 rRNA genes, 22 tRNA genes and a replication control region make up the mitochondrial genome (mt-genome). Within the cell, the mt-DNA genome self-replicates to produce ATP synthase, cytochrome oxidase and the NADH system (Uddin and Chakraborty 2016). The main proteins of the oxidative phosphorylation system are encoded by mitochondrial protein coding genes.

Table 1: Difference between standard genetic code and mitochondrial genetic code

Codon	**'Standard' genetic code**	**'Mitochondrial' genetic codes**			
		Vertebrate	**Invertebrate**	**Yeast**	**Plant**
ATA	Ile	*Met*	*Met*	*Met*	Ile
TGA	STOP	*Trp*	*Trp*	*Trp*	STOP
AGA **AGG**	Arg	*STOP*	*Ser*	Arg	Arg
CTA	Leu	Leu	Leu	*Thr*	Leu
Italics indicate that the code differs from the 'Standard' Code					

The most significant metabolic process occurring in mitochondria is oxidative phosphorylation, wherein an aerobic eukaryotic cell produces ATP using oxygen. There are 13 genes that encode the proteins in the mitochondria and also encode for the protein subunits of the various oxidative phosphorylation complexes. Complex I is made up of 7 mitochondrially encoded proteins (Nd1, Nd2, Nd3, Nd4, Nd4l, Nd5, Nd6), with Nd1 and Nd2 playing a crucial structural function between the complex's peripheral and membrane-embedded arms. Nd2, Nd4 and Nd5 are in charge of transporting electrons (Da Fonseca, Johnson et al. 2008). Complex II is formed by nuclear encoded proteins whereas Complex III is generated by the CYB gene. The generation of a proton gradient through the mitochondrial membrane, which is necessary for ATP synthesis, depends on the CYB gene. As a result, it is crucial in oxidative phosphorylation. Furthermore, MT-CYB gene mutations cause complex III deficit via decreasing oxidative phosphorylation. Cell death is most likely caused by a reduction in the amount of energy available in the cell (Andreu, Checcarelli et al. 2000, Lamantea, Carrara et al. 2002). The mitochondrial respiratory chain is critical in meeting an organism's energy demands (Irwin, Kocher et al. 1991, Degli Esposti, De Vries et al. 1993).

In our study, we examined the nucleotide composition and codon usage pattern of the MT-CYB gene from the Balitoridae family using bioinformatics tools. Additionally, this research would shed light on the processes governing the codon distribution and how they are used differently in the MT-CYB gene, as well as the variables affecting the codon usage pattern of the gene in Balitoridae family.

2. Methodology and materials

2.1. Collection of the coding sequences

The coding sequences (cds) of mitochondrial cytochrome B (MT-CYB) gene were downloaded from the GenBank database in FASTA format maintained by NCBI, USA (https://www.ncbi.nlm.nih.gov/). For *In silico* study, only those cds with valid initiation and termination codons and no uncertain bases were taken into consideration.

2.2. Compositional properties of the CDS

Mitochondrial cytochrome B gene of the Balitoridae family was analysed for the features like: (i) total nucleotide composition and the 3rd codon position nucleotide composition, (ii) total G and C% of synonymous codons in each sequence and also G and C% in the 1st, 2nd and 3rd positions of synonymous codons, (iii) we also estimated different base skews (AT-skew, GC-skews, PU-skew, PY-skew, Ko-skew and Am-skew) to fully comprehend the compositional characteristics of each cds utilising a PERL programme composed by the corresponding author (SC).

2.3. Estimation of synonymous CUB

Below is a discussion of some of the most significant and popular CUB indexes that were examined in this research.

2.3.1. Measure of Effective Number of Codons (ENC)

This parameter is used to calculate the codon usage bias of gene and the value of ENC extends from 20 to 61. The ENC value 20 indicates the existence of the maximum codon bias of a gene indicating that a specific amino acid is codified by only one codon from each family of synonymous codons; whereas, ENC value of 61 denotes that there is minimal or no bias because all of the synonymous codons of each amino acid are used in cds. If the value of ENC of a gene is ≤ 35, then it indicates that the level of CUB is significant (Wright 1990).

To calculate the value of ENC, firstly ‘F_a’ value should be calculated using the formula:

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

Where, the frequency of ‘i’th codon is given by 'p_i', the count of synonymous codons for the 'a'th amino acids is given by k and the observed value of codons for the amino acids is designated by 'na'. The average of the F_a values for every r-fold redundancy class (2, 4, 6-fold for amino acid in translation table 2 of NCBI) was estimated as:

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

Here, ‘*nRC’* denotes the overall amino acids.

The following formula can be used to compute ENC:

$$ENC = \left(\frac{12}{\overline{F_2}}\right) + \left(\frac{6}{\overline{F_4}}\right) + \left(\frac{2}{\overline{F_6}}\right)$$

The symbol ‘F_k’ in the *k*-fold degenerate amino acids stands for the mean of ‘F_k’ values (*k*= 2, 4 and 6 fold degeneracy).

2.3.2. Estimation of Relative Synonymous Codon Usage (RSCU)

This is broadly applied to reveal the codons that occur frequently across genes and genomes. With an assumption that every synonymous member of an amino acid has equal chance of being incorporated in the coding sequence, the RSCU of a codon is measured as the ratio of observed to expected frequency of that codon. RSCU value of a codon greater than 1.0 indicates more frequent usage of the codon than expected and less than 1.0 indicates less frequent usage. RSCU value of more than 1.6 means that the codon is over-represented in coding sequences and less than 0.6 indicates the under-represented codon (Sharp and Li 1986). The following formula was used to measure the RSCU value:

$$RSCUij = \frac{Xij}{\frac{1}{ni}\sum_{j=1}^{ni} Xij}$$

Where "Xij" denotes the frequency of the "jth" codon for the "ith" amino acid and "ni" denotes the number of codons for the "ith" amino acid. Any 'Xij' with a value of zero is given a random value of 0.5.

2.4. Correspondence analysis (COA)

This is used to analyse large-scale CUB heterogeneity patterns and it distributes the codons between two ordinates based on the trends (Shields and Sharp 1987). Each cds is defined as a 60-dimensional vector (for Translation table 2 of NCBI), with each dimension corresponding to the RSCU value of each synonymous codon (total 60 codons) used in the mitochondrial gene. The key trends in codon usage variation may be estimated using relative inertia, which examines the major variables impacting the CUB based on where the genes or genomes are located.

2.5. PR2- bias analysis

PR2 plot (Parity plot) is a scatter diagram and it is drawn with the GC-bias value [G3/ (G3+C3)] as the abscissa and the AT-bias value [A3/ (A3+T3)] as the ordinate (Sueoka 1995). In the centre of the plot both coordinates are 0.5; with A= T and G=C, indicating that there is no bias caused by the two forces of evolution (*i.e.*, mutation and natural selection).

2.6. Neutrality Plot

Strength of evolutionary forces can be determined by using this plot. For this analysis, the GC12 value is plotted vertically, whereas the GC3 value is plotted horizontally in the graph (Sueoka 1988). The regression coefficient (RC) value near to one implies that mutational pressure is dominant over natural selection in shaping CUB; while, natural selection is dominant over mutational pressure when the RC value is close to zero.

2.7. Measure of mutation responsive index (MRI)

In a cds, the degree of mutational drift can be determined from the mutation responsive index value. Mean positive MRI value represent the influence of directional mutation, whereas influence of translational selection is indicated by negative value (Gouy and Gautier 1982).

2.8. Estimation of Translational selection (P2)

It shows the impact of codon-anticodon relationship in determining the translational efficiency of a gene. When the translational selection value is larger than 0.5, it means the coding sequence was favourably selected; and when it is less than 0.5, it means the coding sequence was not favourably selected (Gouy and Gautier 1982).

2.9. Protein properties

We determined the aromaticity, hydrophilicity and GRAVY (grand average of hydropathicity) of the protein encoded by the coding sequence. A hydrophobic protein is represented by a positive GRAVY score, whereas a hydrophilic protein is represented by a negative GRAVY score (Kyte and Doolittle 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.

2.10. Phylogenetic tree analysis of the MT-CYB gene of Balitoridae family

Phylogenetic tree was done by using the MEGA-11 software to analyse the evolutionary relationship of the MT-CYB gene in some species of Balitoridae family.

2.11. Statistical analysis

To estimate the CUB parameters for the MT-CYB gene in freshwater fish family Balitoridae we used a PERL software developed by SC (corresponding author). To determine the relation between each base at the 3rd codon position and the total nucleotide composition, correlation analysis was carried out. The programme "IBM SPSS 21 for Windows" was used to accomplish other statistical analyses.

3. Results and Discussion

3.1. Relation between CUB and expression level

To assess the level of CUB, we calculated the ENC values of the cds for the MT-CYB gene in Balitoridae family. We discovered that the average ENC value in our study was 41.54 i.e., >35. This showed a low CUB of the mitochondrial gene (Butt, Nasrullah et al. 2014).

3.2. Calculation of patterns of codon usage in MT-CYB gene of Balitoridae

The RSCU (Relative synonymous codon usage) value greater than 1.6 indicates the over-represented codon, while RSCU values less than 0.6 indicates an under-represented codon. In Fig.1, we had mentioned the RSCU values of all codons. We found thirteen codons were over-represented and twenty four codons were under-represented in our research [Table 2]. Preferred codons (>1) are listed in Table 3.

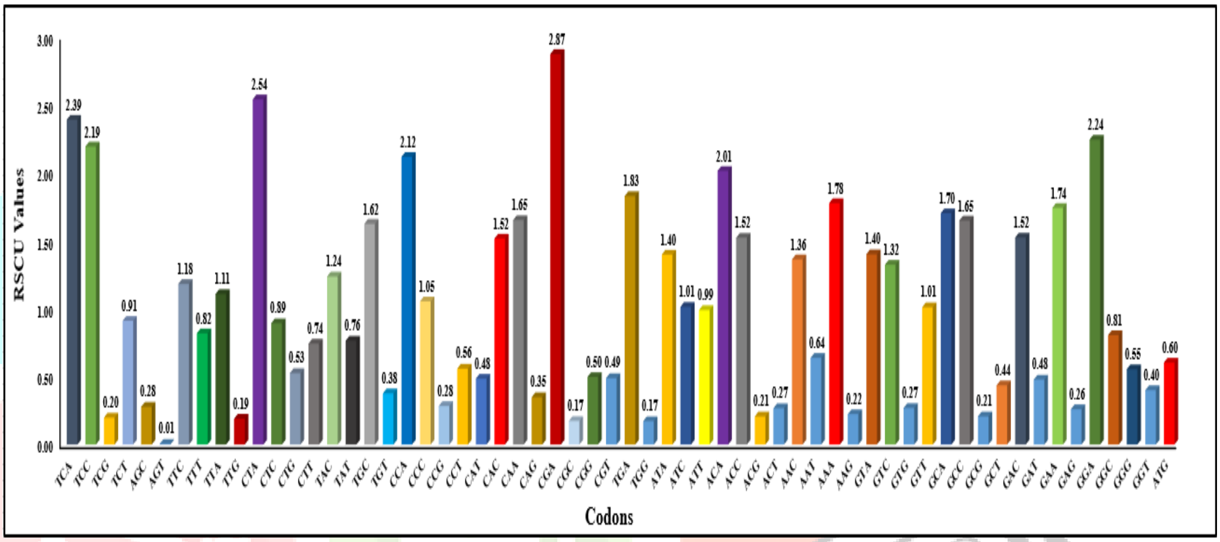


Figure 1: Overall RSCU values of codons in MT-CYB gene of Balitoridae

Table 2: Over-represented and under-represented codons of the MT-CYB gene of *Balitoridae*

No.	Over-represented Codons	Under-represented Codons
1	TCA	TCG
2	TCC	AGC
3	CTA	AGT
4	TGC	TTG
5	CCA	CTG
6	CGA	TGT
7	TGA	CCG
8	ACA	CCT
9	AAA	CAT
10	GCA	CAG
11	GCC	CGC
12	GAA	CGG
13	GGA	CGT
14	-	TGG
15	-	ACG
16	-	ACT

17	-	AAG
18	-	GTG
19	-	GCG
20	-	GCT
21	-	GAT
22	-	GAG
23	-	GGG
24	-	GGT

Table 3: Preferred codons of the MT-CYB gene of *Balitoridae*

Preferred Codons
TCG
AGC
AGT
TTG
CTG
TGT
CCG
CCT
CAT
CAG
CGC
CGG
CGT
TGG
ACG
ACT
AAG
GTG
GCG
GCT
GAT
GAG
GGG
GGT

3.3. Compositional properties of coding sequences

In earlier CUB studies, it was mentioned that the synonymous codon usage of a gene or genome may have been impacted by the total compositional characteristics of the nucleotides (Jenkins and Holmes 2003). In this study, overall nucleotide composition and the base compositions at the 3rd location of codons in the MT-CYB genes of the Balitoridae family were measured (Fig. 2). In this analysis, T (28.56%) and C (28.41%) bases were found in almost equal proportions; while, the proportions of A (27.69%) and G (15.35%) bases were different. The overall GC% content was found to be 43.76%, while the overall AT% was 56.24%, suggesting that the MT-CYB gene was AT-rich (>50%). At the 3rd position of the codons, A (39.10%) was the most common base, followed by C (34.81%), T (19.19%), and G (6.90%). The overall GC3% content was 41.72% and AT3% content was 58.28%; which meant that in the MT-CYB gene, the 3rd position of the codons had abundant AT content than GC content. GC content followed the decreasing trend GC1>GC3>GC2 (Fig. 3).

We found an extremely significant correlation among the ENC and the nucleotide contents of G, C, A, T, G3, C3, A3 and T3 in the MT-CYB gene of the Balitoridae family (*p*<0.01, *p*<0.05). We also noticed a negative correlation between the ENC and the GC content [Table 4]. It suggested that the nucleotide contents had a high impact on CUB in the present study.

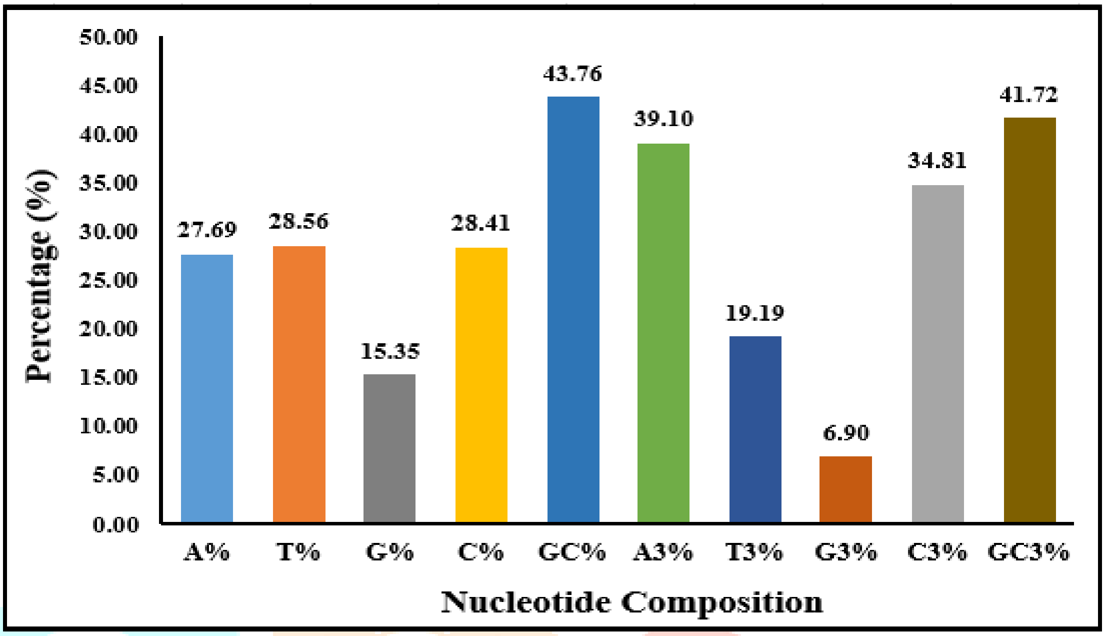


Figure 2: Composition of the nucleotides in the MT-CYB gene of Balitoridae

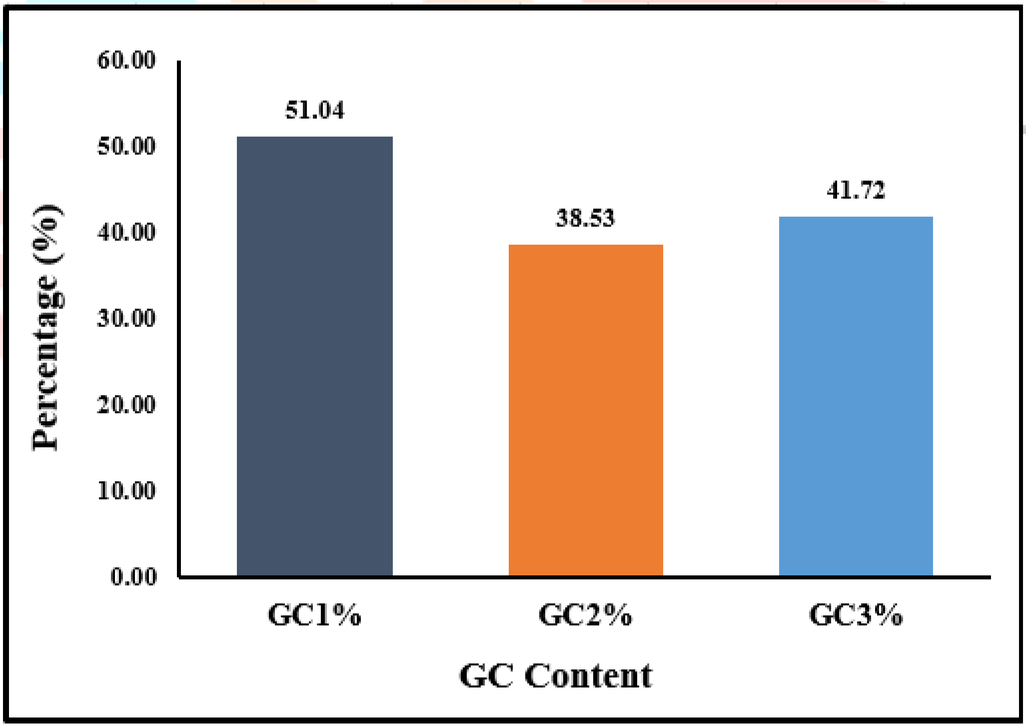


Figure 3: GC contents of MT-CYB gene of Balitoridae

Table 4: Correlation between ENC and base content of the MT-CYB gene of *Balitoridae*

	A%	T%	G%	C%	GC%	A3%	T3%	G3%	C3%	GC3%
ENC	-0.829**	0.809**	0.778**	-0.637**	-0.025	-0.891**	0.828**	0.874**	-.0611**	0.189

**Significant at p < 0.01, *Significant at p <0.05

3.4. Correspondence analysis (COA)

This computational analysis is utilised to estimate the variation of synonymous codon usage patterns. Codons' RSCU values are used for this analysis. In our present study, we used 60 synonymous codons of the MT-CYB gene of Balitoridae (Fig.4). Both the main axes were analysed for contributions to the total variation. In this analysis, axis 1 represented 43.05% of the total variation and axis 2 represented 16.70%. The AT-ending codons were shown in blue dots, whereas the GC-ending codons were shown in red dots in the figure. The patterns of bases around the axes in the figure revealed that mutation pressure might have caused the CUB of the MT-CYB gene during evolution (Wei, He et al. 2014).

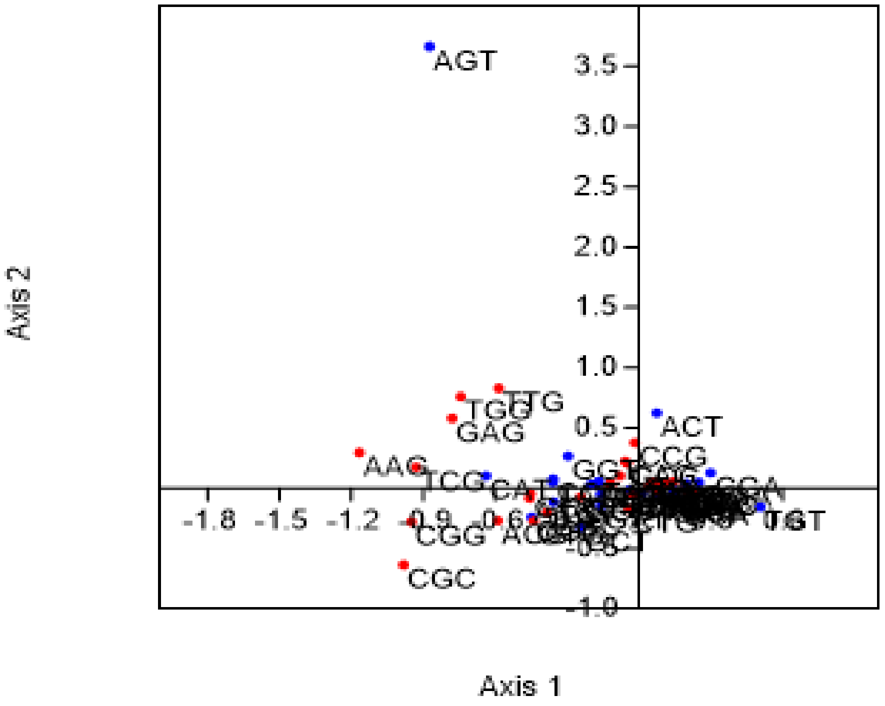
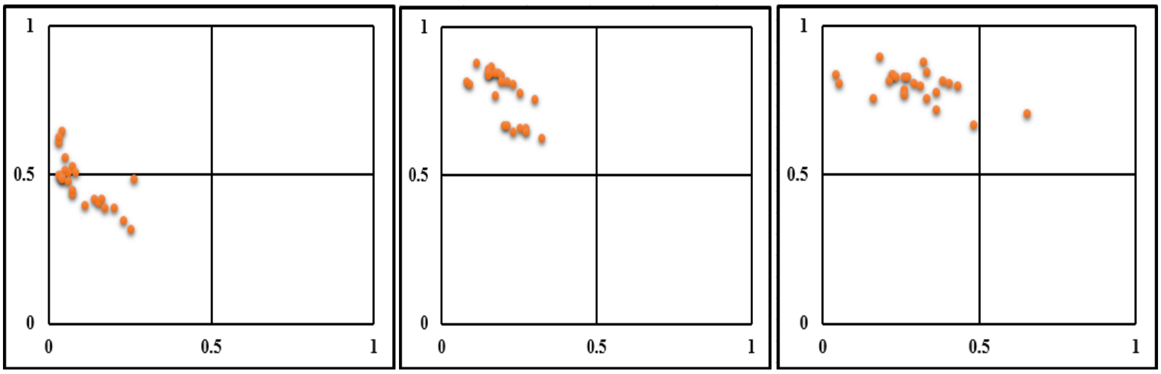


Figure 4: Correspondence analysis (COA) of MT-CYB gene of Balitoridae

3.5. Analysis of the parity rule 2 plot (PR2)

A balanced distribution of bases throughout the PR2 plot explains the impact of mutation on codon usage bias, whereas an asymmetric distribution explains the role of both the evolutionary forces. The investigation was done to determine how evolutionary determinants affected CUB (Fig.5). We plotted (G3/ G3+C3) on the horizontal axis, whereas (A3/ A3+T3) on the vertical axis. Across the plot we detected an unequal distribution of bases, showing that both evolutionary forces impacted the CUB of MT-CYB gene in Balitoridae (Deb, Uddin et al. 2018).



(A) (B) (C)

Figure 5: Parity bias plots (PR2) analysis showing (A) 2-fold, (B) 4-fold and (C) 6-fold degenerate codons in MT-CYB gene of Balitoridae

3.6. Analysis of the neutrality plot

This parameter is used to assess how evolutionary forces and the relationship between GC3 and GC12 have affected the evolution of mitochondrial gene. It is the regression of GC12 on GC3. In our study, a positive correlation between GC3 and GC12 (r = 0.304, $p<0.01$) was found (Sueoka 1988). Furthermore, we produced a linear regression graph by plotting GC3 on the X-axis and GC12 on the Y-axis (Fig. 6). For the MT-CYB gene of the Balitoridae family, the regression coefficient (RC) between GC12 and GC3 was found 0.0768, which meant that the relative neutrality for GC3 was close to 7.7%, and the relative constraint was close to 92.3%. With a ratio of 0.0768/0.9232=0.08, both evolutionary factors had an impact on the GC12 in MT-CYB gene of Balitoridae family. When the RC value < 0.5, natural selection is considered to be more important but mutational pressure is thought to be more important when the RC value > 0.5. From these findings, we could infer that natural selection was dominant over the mutational pressure in controlling the CUB of the MT-CYB gene of Balitoridae family.

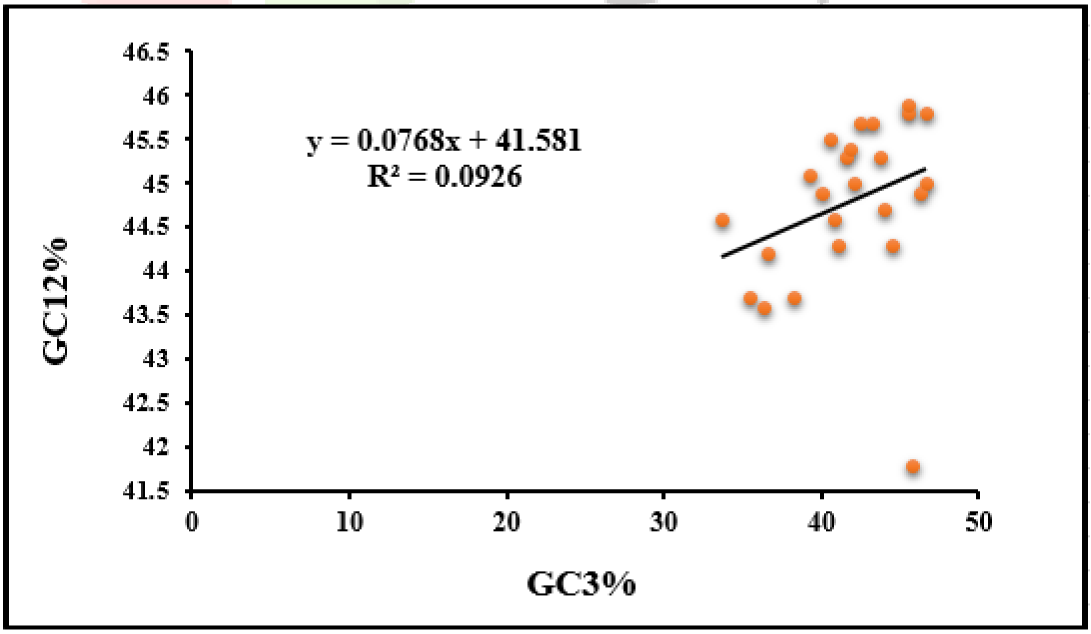


Figure 6: Neutrality plot analysis of MT-CYB gene of Balitoridae

3.7. Interrelationships among base compositions

The variance in codon usage patterns is mostly due to two evolutionary processes (Mazumder, Uddin et al. 2018, Mazumder, Uddin et al. 2018). Correlation analysis of compositional properties (Chen 2013) was used to determine the primary forces that influence the display of CUB in the gene. Using Karl Pearson's method, we compared the nucleotide contents of third codon position with the overall nucleotide contents [Table 5]. A highly significant correlation (at $p<0.01$, 0.05) was observed, which implied that mutation pressure might be the main deciding force in the CUB of MT-CYB gene of Balitoridae family (Zhao, Zhang et al. 2007, Zhang, Dai et al. 2013).

Table 5: Correlation coefficients between overall nucleotide contents and nucleotide contents in the third position of codons in MT-CYB gene of *Balitoridae*

	A3%	**T3%**	**C3%**	**G3%**	**GC3%**
A	0.924**	-0.586**	-0.811**	0.219	-0.501*
T	-0.594**	0.988**	0.549**	-0.896**	-0.360
G	-0.830**	0.496*	0.801**	-0.229	0.483*
C	0.325	-0.888**	-0.374	0.975**	0.585**
GC	-0.314	-0.477*	0.243	0.763**	0.924**

**Significant at p < 0.01, *Significant at p <0.05

3.8. Role of translational selection (P2)

This parameter is used to determine whether translational selection has any impact on the mitochondrial gene under study. In our research, we noticed the mean P2 value as 0.51, i.e., P2 was >0.50. The P2 value >0.50 suggested that the MT-CYB gene was more significantly affected by translational selection (Chakraborty, Deb et al. 2019). Correlation between P2 and ENC values in our study showed a negative, highly significant correlation (r = -0.656**, p<0.01), which suggested that ENC and translational selection had the opposite relationships (Deb, Uddin et al. 2020). As a result, it was inferred that the coding sequences of MT-CYB gene with low codon bias (as revealed by high ENC values) might have been the under the influence of low translational selection during the course of evolution.

3.9. Nucleotide skewness analysis

In our study, the mean values of AT-skew and GC-skew were -0.02 and -0.30 respectively, for MT-CYB gene of Balitoridae family. The MT-CYB gene clearly indicated that T and C nucleotides were used more often than A and G (Wei, He et al. 2014). Previous research on the CUB of genes demonstrated that base skewness played a significant impact in determining the codon usage bias (Deb, Uddin et al. 2018). In the present study, Karl Pearson's correlation method was used to compare base skewness values with ENC and we found a positive, highly significant correlation of ENC with GC-skew (0.877) and PY-skew (0.738); whereas a negative, highly significant correlation with PU-skew (-0.809) and AT-skew (-0.922). We also found a negative correlation of ENC with Am-skew (-0.267) and Ko-skew (-0.154). These indicated that the CUB of MT-CYB gene of the Balitoridae family might have been affected by the nucleotide skewness in cds.

3.10. Correlation analysis of CUB with protein properties

Pearson's correlation analysis was applied to investigate the relationship of CUB with GRAVY, aromaticity and hydrophilicity of the encoded protein for the MT-CYB gene. It was evident from this research that the protein aromaticity displayed a highly significant positive correlation (0.568**, $p<0.01$) with ENC, indicating that protein’s aromaticity values were related to CUB in the MT-CYB gene.

3.11. Mutation Responsive Index (MRI)

MRI is employed to determine how translational selection and mutation could affect the CUB of genes. The role of mutation pressure is denoted by a positive MRI value, while translational selection is denoted by a negative MRI value. In our research, the average MRI value was found as -0.70, which indicated that the MT-CYB gene of Balitoridae family showed the greater effect of translational selection.

3.12. Phylogenetic analysis of MT-CYB gene in some species of Balitoridae

Phylogenetic clustering analysis using MEGA 11 software was done for the MT-CYB gene in some species of Balitoridae fish family (Fig. 7). We had noticed that the MT-CYB gene of almost all the species of Balitoridae family in our study diverged from a common ancestral gene, except the gene in the mitochondrion of *Homaloptera confuzona* voucher CBMZF 11705, which had a different origin from the rest of the species.

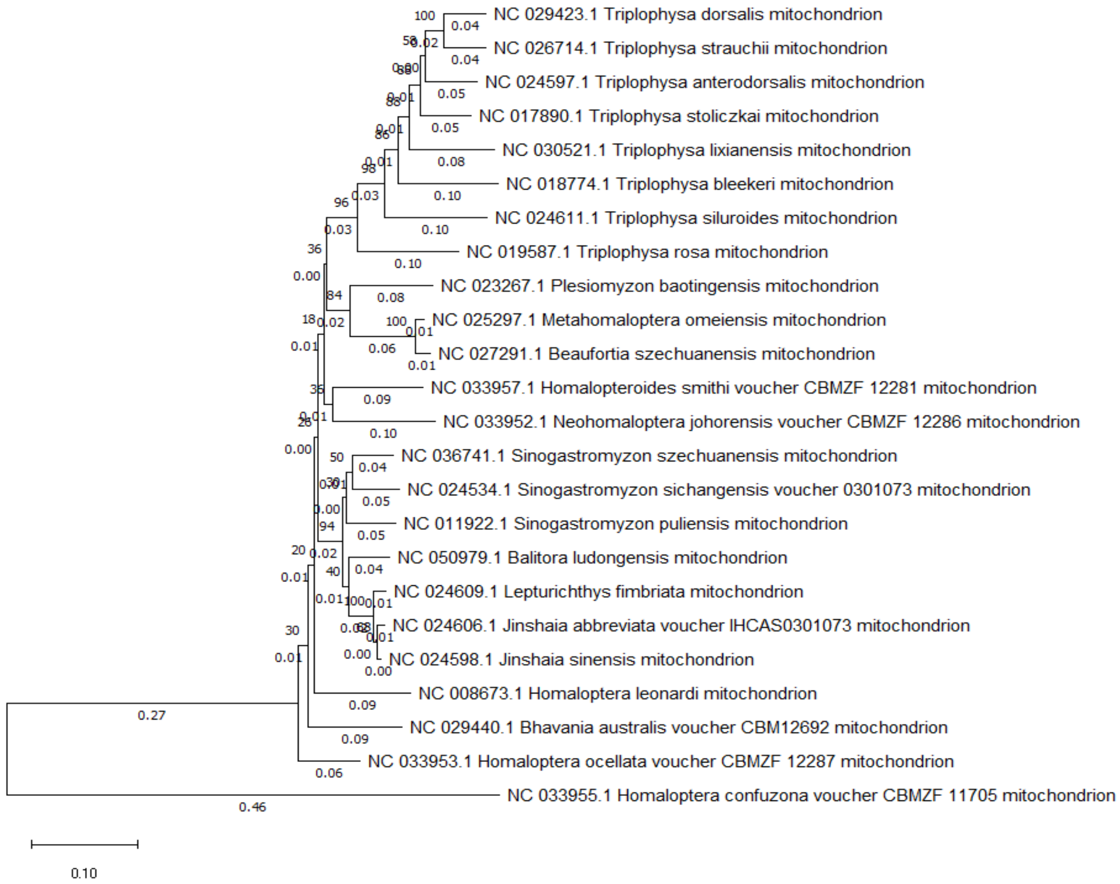


Figure 7: Phylogenetic tree analysis of the MT-CYB gene of some species of Balitoridae

4. Discussion

The preferential use of some synonymous codons over others for the same amino acid in RNA transcripts is known as codon usage bias (CUB) (Baba, Ara et al. 2006). The final gene product may be strongly impacted by this nonrandom pattern of codon selection. Many tRNA molecules are thought to be present in rapidly growing organisms, which increases the likelihood that the highly expressed genes in the species will have a high level of codon usage bias (Rocha 2004). In adaptive and evolutionary biology, CUB is a crucial component. Additionally, by optimizing codons, CUB helps to enhance the expression of foreign genes (Deb, Uddin et al. 2021). This study looked at the codon usage pattern and the factors influencing the MT-CYB gene in the Balitoridae family.

In our study we found that the average ENC value was 41.54, which is greater than 35. This indicates a low codon usage bias in the mitochondrial gene. A low codon usage bias (CUB) means that almost all synonymous codons were used to encode an amino acid, with little or no preference for certain codons. Furthermore, low CUB might make it easier for diverse cell types with different codon preferences to replicate. Uddin et al. (2016) reported in their study that the average ENC values for the MT-CYB gene in several species of aves, pisces and mammals respectively were -59.66 ± 0.61, 58.33 ± 2.94 and 58.33 ± 2.12. The high ENC value suggested that the CUB for the MT-CYB gene was low and that it was maintained almost at the same level across the species under investigation (Uddin and Chakraborty 2016).

According to the RSCU values, we found twenty-four under-represented codons and thirteen over-represented codons in the Balitoridae family's mitochondrial cytochrome B gene. Uddin et al. (2016) in their RSCU analysis of the MT-CYB gene of pisces, aves and mammals found that at the 3rd codon location, A and C nucleotides were most frequently used than T and G. While taking the mean RSCU values, they also observed that 27, 29 and 32 codons in aves, mammals and pisces were the more frequently used ones among the 60 sense codons. The RSCU values of 60 sense codons confirmed once more that the MT-CYB gene has a weak CUB (Uddin and Chakraborty 2016).

The nucleotide composition of a gene determines the pattern of codon usage. Although the bases A (27.69%) and G (15.35%) had different proportions in our analysis, the bases T (28.56%) and C (28.41%) were present in almost equal percentages. With an overall GC content of 43.76% and an overall AT content of 56.24%, it can be concluded that the MT-CYB gene is AT-rich (above 50%).Uddin et al. (2016) in their study found that the distributions of the nucleotides (T, A, C and G%) among the codons vary in different species of aves, mammals and pisces; C/A ending codons were more preferred in aves, T/C ending codons in pisces and A/C ending codons in mammals. They also noticed GC content was lesser than AT content (Uddin and Chakraborty 2016).

The trend of codon usage variation in the mitochondrial CYB gene of Balitoridae family was identified by correspondence analysis. Furthermore, we observed that the distribution of AT and GC ending codons were both closer and farther from the axes, indicating the possibility that CUB was guided by natural selection as well as mutation pressure. In their study, Uddin et al. (2016) discovered the uneven distribution of AT and

GC ending codons across the axes. They found that the first axis occupied 34.53% and the second axis occupied 12.79% of the overall variation in the MT-CYB gene (Uddin and Chakraborty 2016).

In the parity rule 2 plot (PR2) analysis of our study, the variance in base distribution across 2-fold, 4-fold, and 6-fold degenerate codons was shown. Additionally, it demonstrated how natural selection and mutation work together to determine the CUB throughout the MT-CYB gene of the Balitoridae family. In their investigation, Parvin et al. (2020) discovered that in all orders of amphibia, the nucleotides A and T were not proportional to G and C; this reinforced the idea that, both the evolutionary forces might have shaped the codon usage bias of the cytochrome B gene (Barbhuiya, Uddin et al. 2020).

The effect of evolutionary forces. i.e., natural selection and mutation pressure, was examined using the neutrality plot analysis. Our research revealed that in the MT-CYB gene of Balitoridae family, natural selection was more dominant than mutational pressure. Parvin et al. (2020) found that the RC values of GC12 on GC3 in urodela, gymnophiona and anura respectively were 0.1439, 0.1704 and 0.1061. Since the RC values were close to zero, natural selection was thought to be to be the prime force in shaping the codon usage bias of MT-CYB genes of those amphibian species (Barbhuiya, Uddin et al. 2020).

We found highly significant correlation between compositional constraints by correlation analysis, suggesting that mutation pressure could be the primary determining factor in the codon usage bias (CUB) of the MT-CYB gene in the Balitoridae family. In their study, Parvin et al. (2020) found a strong significant correlation among the total base composition and the nucleotide composition at the 3rd codon position for homogeneous combinations. With the exception of A & T3, T & A3 in anura, and A & T3 in gymnophiona; the overall nucleotide and the base composition at the 3rd codon location of the codons for heterogeneous combinations also showed significant correlation in majority of the cases as well. This study implied that both the forces were involved in determining the codon usage bias of mitochondrial cytochrome B gene of amphibians across all the orders (Barbhuiya, Uddin et al. 2020).

In our study, we discovered that the MT-CYB gene showed negative values for both AT and GC skew. This indicated a clear preference for T and C nucleotides over A and G nucleotides. In their study, Uddin et al. (2016) discovered negative values of GC-skews in some species; while positive values of AT-skews in the majority of species of aves, pisces and mammals, indicating that C was more abundant than G in most species and A was more numerous than T in the others (Uddin and Chakraborty 2016).

Our analysis showed a highly significant positive correlation between protein aromaticity and ENC, indicating a strong relationship between protein aromaticity and codon usage bias (CUB) in the MT-CYB gene.In their work, Uddin et al. (2016) discovered a significant negative correlation between the aromaticity of the protein and the ENC values, which shows that a reduction in aromatic amino acids in the protein could increase the ENC of the gene (Uddin and Chakraborty 2016).

5. Conclusion

In this work, the composition of the nucleotides and the CUB of the protein-coding gene MT-CYB of the Balitoridae family were examined. The CUB of the MT-CYB gene was low, which may indicate that there was a great variability in synonymous codon usage of the gene. The base frequencies of T and C were almost similar in the gene, but G and C had different frequencies. The Balitoridae family had an AT-rich MT-CYB gene. Thirteen codons were found to be overrepresented while twenty four codons were found to be underrepresented in our study. Both evolutionary forces appear to have had an impact on the CUB of mitochondrial cytochrome B gene during the evolution of the Balitoridae fish family. Mutation pressure had a negligible impact, whereas natural selection had the predominant effect on the evolution of this gene. The current study advances our knowledge of the mechanisms influencing the codon usage as well as the evolutionary progress of the MT-CYB gene in the Balitoridae family.

6. Funding

The work was unfunded.

7. Compliance with Ethical Standards

I. Conflict of interests

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

II. Research involving Human Participants and/or Animals

Not Applicable. This study is based on DNA sequence analysis and the data were retrieved from publicly available databases.

III. Consent for publication

JCD- Original Draft, Data curation, Analysis of data, Investigation, Interpretation, Validation.

SC- Conceptualization, Project administration, Formal analysis, Software, Visualization, Supervision, Methodology, Writing- Review and Editing

8. Acknowledgments

The authors are thankful to Assam University, Silchar 788011, Assam, India for providing the necessary facilities for this work.

9. References

Akashi, H. (1997). "Codon bias evolution in Drosophila. Population genetics of mutation-selection drift." Gene **205**(1-2): 269-278.

Andreu, A. L., et al. (2000). "A missense mutation in the mitochondrial cytochrome b gene in a revisited case with histiocytoid cardiomyopathy." Pediatric research **48**(3): 311-314.

Baba, T., et al. (2006). "Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection." Molecular systems biology **2**(1): 2006.0008.

Barbhuiya, P. A., et al. (2020). "Analysis of compositional properties and codon usage bias of mitochondrial CYB gene in anura, urodela and gymnophiona." Gene **751**: 144762.

Behura, S. K. and D. W. Severson (2012). "Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes."

Bulmer, M. (1991). "The selection-mutation-drift theory of synonymous codon usage." Genetics **129**(3): 897-907.

Butt, A. M., et al. (2014). "Genome-wide analysis of codon usage and influencing factors in chikungunya viruses." PloS one **9**(3): e90905.

Chakraborty, S., et al. (2019). "Analysis of codon usage patterns and influencing factors in Nipah virus." Virus research **263**: 129-138.

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**.

Da Fonseca, R. R., et al. (2008). "The adaptive evolution of the mammalian mitochondrial genome." BMC genomics **9**(1): 1-22.

Deb, B., et al. (2020). "Codon usage pattern and its influencing factors in different genomes of hepadnaviruses." Archives of virology **165**(3): 557-570.

Deb, B., et al. (2021). "Analysis of codon usage of Horseshoe Bat Hepatitis B virus and its host." Virology **561**: 69-79.

Deb, B., et al. (2018). "Analysis of codon usage pattern of mitochondrial protein-coding genes in different hookworms." Molecular and biochemical parasitology **219**: 24-32.

Degli Esposti, M., et al. (1993). "Mitochondrial cytochrome b: evolution and structure of the protein." Biochimica et Biophysica Acta (BBA)-Bioenergetics **1143**(3): 243-271.

Gouy, M. and C. Gautier (1982). "Codon usage in bacteria: correlation with gene expressivity." Nucleic acids research **10**(22): 7055-7074.

Hershberg, R. and D. A. Petrov (2008). "Selection on codon bias." Annual review of genetics **42**: 287-299.

Irwin, D. M., et al. (1991). "Evolution of the cytochromeb gene of mammals." Journal of molecular evolution **32**(2): 128-144.

Jenkins, G. M. and E. C. Holmes (2003). "The extent of codon usage bias in human RNA viruses and its evolutionary origin." Virus research **92**(1): 1-7.

Jermiin, L. S., et al. (1994). "Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome b genes." Journal of molecular evolution **39**(2): 160-173.

Jukes, T. H. and V. Bhushan (1986). "Silent nucleotide substitutions and G+ C content of some mitochondrial and bacterial genes." Journal of molecular evolution **24**(1): 39-44.

Kottelat, M. (2012). "Conspectus cobitidum: an inventory of the loaches of the world (Teleostei: Cypriniformes: Cobitoidei)." Raffles Bulletin of Zoology.

Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." Journal of molecular biology **157**(1): 105-132.

Lamantea, E., et al. (2002). "A novel nonsense mutation (Q352X) in the mitochondrial cytochrome b gene associated with a combined deficiency of complexes I and III." Neuromuscular disorders **12**(1): 49-52.

Lobry, J. and C. Gautier (1994). "Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 Escherichia coli chromosome-encoded genes." Nucleic acids research **22**(15): 3174-3180.

Mazumder, G. A., et al. (2018). "Codon usage pattern of complex III gene of respiratory chain among platyhelminths." Infection, Genetics and Evolution **57**: 128-137.

Mazumder, G. A., et al. (2018). "Preference of A/T ending codons in mitochondrial ATP6 gene under phylum Platyhelminthes: codon usage of ATP6 gene in Platyhelminthes." Molecular and biochemical parasitology **225**: 15-26.

Muto, A. and S. Osawa (1987). "The guanine and cytosine content of genomic DNA and bacterial evolution." Proceedings of the National Academy of Sciences **84**(1): 166-169.

Paxton, J. R. and W. N. Eschmeyer (1998). "Encyclopedia of fishes."

Randall, Z. S. and L. M. Page (2015). "On the paraphyly of Homaloptera (Teleostei: Balitoridae) and description of a new genus of hillstream loaches from the Western Ghats of India." Zootaxa **3926**(1): 57-86.

Rocha, E. P. (2004). "Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization." Genome research **14**(11): 2279-2286.

Sauna, Z. E. and C. Kimchi-Sarfaty (2011). "Understanding the contribution of synonymous mutations to human disease." Nature Reviews Genetics **12**(10): 683-691.

Schutz, F. A., et al. (2013). "Single nucleotide polymorphisms and risk of recurrence of renal-cell carcinoma: a cohort study." The lancet oncology **14**(1): 81-87.

Seligmann, H. and G. Warthi (2017). "Genetic code optimization for cotranslational protein folding: codon directional asymmetry correlates with antiparallel betasheets, tRNA synthetase classes." Computational and Structural Biotechnology Journal **15**: 412-424.

Shackelton, L. A., et al. (2006). "Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses." Journal of molecular evolution **62**(5): 551-563.

Shang, M., et al. (2011). "Analysis on codon usage of chloroplast genome of Gossypium hirsutum." Scientia Agricultura Sinica **44**(2): 245-253.

Sharp, P. M. and W.-H. Li (1986). "An evolutionary perspective on synonymous codon usage in unicellular organisms." Journal of molecular evolution **24**(1-2): 28-38.

Sharp, P. M. and G. Matassi (1994). "Codon usage and genome evolution." Current opinion in genetics & development **4**(6): 851-860.

Shields, D. C. and P. M. Sharp (1987). "Synonymous codon usage in Bacillus subtilis reflects both translational selection and mutational biases." Nucleic acids research **15**(19): 8023-8040.

Stenico, M., et al. (1994). "Codon usage in Caenorhabditis elegans: delineation of translational selection and mutational biases." Nucleic acids research **22**(13): 2437-2446.

Sueoka, N. (1988). "Directional mutation pressure and neutral molecular evolution." Proceedings of the National Academy of Sciences **85**(8): 2653-2657.

Sueoka, N. (1995). "Intrastrand parity rules of DNA base composition and usage biases of synonymous codons." Journal of molecular evolution **40**(3): 318-325.

Uddin, A. and S. Chakraborty (2016). "Codon usage trend in mitochondrial CYB gene." Gene **586**(1): 105-114.

Wei, L., et al. (2014). "Analysis of codon usage bias of mitochondrial genome in Bombyx mori and its relation to evolution." BMC evolutionary biology **14**(1): 1-12.

Wright, F. (1990). "The 'effective number of codons' used in a gene." Gene **87**(1): 23-29.

Zhang, Z., et al. (2013). "Synonymous codon usage in TTSuV2: analysis and comparison with TTSuV1." PloS one **8**(11): e81469.

Zhao, S., et al. (2007). "The factors shaping synonymous codon usage in the genome of Burkholderia mallei." Journal of Genetics and Genomics **34**(4): 362-372.

Zhou, T., et al. (2005). "Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses." Biosystems **81**(1): 77-86.

List of Figures:

List of Tables: