Original Research

Comparative studies of codon usage profile of *Anisakis simplex* (Nematoda) and *Carassius gibelio* (Prussian carp)

W. Ahmed[1,2], S. Gupta[1], I. Mukherjee[3], V. Babu[1,2] and R. Singh[1,2,]*

[1]CSIR-Indian Institute of Integrative Medicine Jammu- 180 001, India
[2]Academy of Scientific and Innovative Research (AcSIR), Jammu- 180 001, India
[3]Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, Na Sádkách, České Budejovice-11720, Czech Republic

***Corresponding Author Email :** rubail2010@gmail.com
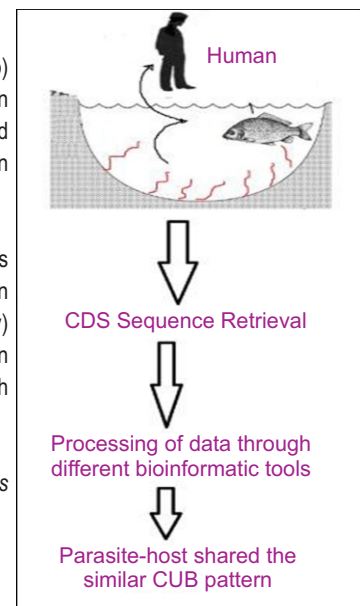
Abstract

Aim: The aim of the present study was to understand the molecular relationship between nematode (parasite) and fish (host) through codon usage bias (CUB) analysis.

Methodology: The Codon usage bias analysis has been performed in fish *Carassius gibelio* (Prussian carp) and nematode fish parasite *Anisakis simplex*. The complete coding sequences (CDS) of *C. gibelio* (Prussian carp) and *A. simplex* (Nematode) were retrieved from National Center for Biotechnology Information and followed to that we have performed bioinformatics analysis to understand the codon usage pattern between host and parasite.

Results: Different CUB indices like Relative synonymous codon usage (RSCU), Effective number of codons (ENC), Codon adaptation index (CAI) and Codon bias index (CBI) revealed a similar pattern in the codon usage in *C. gibelio* and *A. simplex*. In addition, inclusive analysis using different plots (ENC, parity, neutrality) had shown the influence of both the evolutionary forces i.e mutational and translational selection on codon usage pattern. This describes the role of evolutionary forces in determining the conserved genome to establish species-specific function-level differences for efficient survival.

Interpretation: The present study elucidated the association between *Carassiusgibelio* (host) and *Anisakis simplex* (parasite) based on the similar pattern of codon usage bias between both the species.

Key words: *Anisakis simplex, Carassius gibelio*, Codon usage bias, Mutations, Natural selection, Parasite

Human

CDS Sequence Retrieval

Processing of data through different bioinformatic tools

Parasite-host shared the similar CUB pattern

Introduction

Codon usage bias (CUB) refer to the differential usage frequency of synonymous codons influenced by different factors on genes during the course of evolution (Frumkin *et al.*, 2018). It is said that codon usage not only indicate the origin, evolution and mutation mode of species or genes, but also has an effect on gene function and protein expression (Pop *et al.,* 2014; Quax *et al.,* 2015; Tuller *et al.,* 2010). This phenomenon commonly reported in numerous organisms from prokaryotes to eukaryotes and this codon usage biasness is generally found in highly expressed genes (Akashi, 1994). The genetic code comprised of 64 codons encoded by 20 amino acids. Therefore, some amino acids are encoded by more than one codon, they are known as synonymous codons (Mitra *et al.*, 2016). Among the synonymous codons there is difference in usage frequency, meaning that some codons are rarely used while other codons are frequently used in a particular organism.

Different factors like transcription/translation rate, tRNAabundance, gene length, RNA secondary structure, protein structure, hydrophobicity, GC content etc influence the codon usage bias in organisms (Gupta *et al.,* 2004; Ikemura, 1981; Kahali *et al.,* 2008; Martín *et al.,* 1989; Moriyama and Powell, 1998; Romero *et al.,* 2000; Sharp and Li, 1986). Further all these factors are influenced by two evolutionary forces one is mutation/non-random usage of synonymous codon and another is natural selection/fitness (Gupta and Ghosh, 2001; Mougel *et al.,* 2004). CUB analysis helps to understand the evolution and environmental adaptation of living organisms (Angellotti *et al.,* 2007). In addition it also provides a lot of information regarding parasite and host functional relationship. The evolutionary changes that allow the parasite to better adapt towards the environmental conditions and their host can be reflected by the patterns of CUB (Biswas *et al.,* 2019).

The co-evolution and adaptation of parasites to the hosts are generally studied by analyzing the synonymous CUB in the complete coding sequence and the specific functional genes (Cutter *et al.,* 2006; Duret, 2002; Liu *et al.,* 2010; Wong *et al.,* 2010; Zhao *et al.,* 2003). Nematode parasites of aquatic origin are profoundly widespread and diverse in freshwater and marine ecosystems (Moravec, 1998). Aquatic nematodes can be free living or found as a parasite in fishes, crustaceans, squid and marine vertebrates and are cosmopolitan in dispersion (Bullini 1997). One of the commonly found parasitic nematode is *Anisakis,* which is transmissible to humans through consuming raw fish. Gibel carp (*Carassius gibelio*) is widespread across Europe and also in North and East Asia. It inhabits in wide varieties of water resources and lowland rivers, usually allied with submerged vegetation or regular flooding (Kottelat and Freyhof, 2007). *C. gibelio* is considered as one of the most harmful fish species for native fish communities (Crivelli, 1995). Different type of parasitic infections in this species leads to decrease in the population of all native fish species, fish eggs, invertebrates, plants and all animals in the ecosystem. Therefore, since both *A. simplex* and *C. gibelio* are extremely important in terms of ecological and societal point of view, the present study conducted to explore the codon usage pattern between *C. gibelio* and *A. simplex* and to understand its impact on host-parasite relationship. Moreover, since the key data on the codon usage bias impacting fish-nematode interaction is limited, the current investigation will help in understanding the role of codon usage pattern with respect to the prevalence of nematode parasites towards their host.

Materials and Methods

Sequence data retrieval: The complete CDS of *C. gibelio* (Prussian carp) and *A. simplex* (Nematode) were retrieved from National Centre for Biotechnology Information (https://www.ncbi. nlm.nih.gov/). The correct and complete coding sequences with appropriate start codon and stop codons are analysed.

Relative synonymous codon usage (RSCU) analysis: RSCU is the ratio of the observed frequency of a codon to the expected frequency under random codon usage. RSCU value>1.0 indicates that the corresponding codon was used more frequently than the expected frequency while RSCU values < 1.0 indicates that the particular codon was used less frequently (Jenkins and Holme, 2003). RSCU value for each codon was estimated as per Sharp and Li `(1986).

Effective number of codons (ENC): Effective number of codons (ENC) is the most extensively used factor to measure the codon usage bias (Sharma *et al.,* 2014). ENC value ranges from 20-61. ENC value less than 35 represents high codon usage bias whereas higher ENC value (greater than 35) represents low codon usage bias. ENC was calculated as per the formula given by Wright (1990).

Neutrality plot: Neutrality plot is primarily used to assess the influence of mutational pressure on codon usage bias. In neutrality plot, GC3 and GC12 (mean GC1 and GC2) values are plotted on the lateral and vertical axis, respectively to produce a scatter diagram. The slope of the regression line in neutrality plot close to 1 indicates the dominance of mutation pressure while deviation from 1 indicates the effects of other forces such as selection constraints (Sueoka, 1988).

PR2-bias plot: Parity rule 2 (PR2) plot was constructed to evaluate the influence of evolutionary forces(mutation pressure and natural selection) on the codon usage. The centre of the plot, where both coordinates were 0.5, is the place where A = T and G = C and it indicates that there is no bias, whereas deviation from the centre point indicates the role of mutation or selection between two DNA strands. The A3s/(A3s+T3s) and G3s/(G3s+C3s) of each gene plotted as the ordinate and the abscissa to explore the relationship between purines (A and G) and pyrimidines (T and C) at the third codon position (Sueoka, 1995).

Correspondence analysis (COA): Correspondence analysis is a multivariate statistical technique (Greenacre, 1984), where all

genes are plotted in a 59-dimensional hyperspace in accordance to their usage of 59 sense codons. The RSCU score was used as an input in Past 4.3 program to get the 59 axis coordinates. Subsequently, the two major axes were identified representing most of the variation in CUB.

Softwares used: CodonW and INCA2.0 was used to analyse nucleotide composition and different codon usage indices. The values of GC1, GC2, and GC3 were obtained from computational bioinformatics analysis (http:// agnigarh.tezu.ernet.in/~ssankar/cub.php). Past 4.3 was used for COA. Statistical analysis was performed using SPSS 16.0.

Results and Discussion

Codon usage bias is an important criterion that significantly varies among different organisms and helps us to understand the molecular organization and the evolutionary pattern of genomes (Deb *et al.,* 2020). Previously, the genomic studies have revealed that the codon usage pattern of parasite may have acquired influence from its host to some extent. Many studies were carried out to consider this assumption in viruses and bacteria (Butt *et al.,* 2014; Sahu *et al.,* 2005). The present study was performed to compare the codon usage pattern in fish *C. gibelio* and a nematode fish parasite, *A. simplex. A. simplex* is a well-known parasite of carp fish, although it is still not reported in *C. gibelio*. However, the present investigation will help us to explore the ecological relationship among these two important species. The main emphasis was on the role of CUB on host-parasite interaction in addition to the role of evolutionary forces in shaping the codon pattern. Codon usage bias can be significantly affected by the overall nucleotide composition of the genome (Jenkins and Holmes, 2003; Nasrullah *et al.,* 2015).

Nucleotide composition analysis showed that in *C. gibelio* the mean percentage of C was found to be highest followed by the mean percentage of G, T and A whereas in *A. simplex*, the mean percentage of A was found to be greater than T > C > G. The analysis of GC content at three codon positions viz GC1, GC2, and GC3 revealed that mean GC at the second position was maximum in both *C. gibelio* and *A. simplex*. Mean GC3 content was less than the mean AT3 in both the host and pathogen.

Therefore, nucleotide composition analysis revealed that the coding sequences of *C. gibelio* and *A. simplex* are AT biased (Table1). It has been reported that GC rich organisms tend to prefer GC rich codons whereas low GC content organisms favour AT preferred codons, suggesting the role of evolutionary pressures acting in the same direction as nucleotide substitution biases that determine overall nucleotide content of genomes (Hershberg and Petrov 2009, Kawabe and Miyashita 2003). Codon usage pattern of *C. gibelio* and its parasite *A. simplex* were evaluated using different indices like Effective number of Codons (ENC), Codon Adaptation Index (CAI) and Codon bias index (CBI). The ENC value for *C. gibelio* and *A. simplex* was found to be greater than 35, which indicates the random usage of codons and weak codon usage bias and low gene expression. It was found that highly expressed genes exhibited greater codon bias than lowly expressed genes (Sahoo *et al.,* 2019). CAI and CBI are considered as universal measure of codon bias and their value ranges from 0 to 1, where 0 indicates the random codon usage and 1 signifies the preferred codon usage.

It has been found that the value of CBI and CAI for *C. gibelio* and *A. simplex* was less than 1, further reflecting weak codon usage and low gene expression (Kiewitz and Tümmler, 2000). Overall, the results of all three indices indicate weak codon usage bias in *C. gibelio* and *A. simplex*. Relative synonymous codon usage (RSCU) analysis was carried out to identify the preferred synonymous codon for each amino acid. Based on the frequency, it was found that *C. gibelio* and *A. simplex* tend to show similarity in the use of preferred codons with respect to nine amino acids, namely Ala (GCU), Asp (GAU), Phe (UUC), Ile (AUC), Lys (AAA), Asn (AAC), Pro (CCA), Thr (ACA), and Tyr (UAC). These uniformities suggest the influence of selection pressure on codon usage bias and further the ability of the parasite to acclimatize itself inside the host's environment (Butt *et al.,* 2014; Ma *et al.,* 2015; Wong *et al.,* 2010).

Both *C. gibelio* and *A. simplex* preferred AC ending codons at third codon position exhibiting the role of both mutational pressure and selection in shaping the codon usage pattern (Hershberg and Petrov, 2009). The RSCU values for each synonymous codon from both the species are shown in Table 2. ENC plot analysis was carried out in order to determine the factors

Table 1: Nucleotide composition analysis of *Carassius gibelio* and *Anisakis simplex*

Nucleotide composition	Carassius gibelio	Anisakis simplex
A3	28.2 %	35.8 %
T3	31.3 %	34.2 %
C3	33.7 %	33.2 %
G3	32.2 %	25.7 %
GC	49.78 %	48.10 %
GC1	49.54%	47.79%
GC2	52.41%	59.03%
GC3	42.76%	38.49%
AT3	57.23%	61.5%
Gc12	75.75%	77.31%

Table 2: RSCU values analysis of *Carassius gibelio* and *Anisakis simplex*. The shared codons between *C. gibelio* and *A. simplex* are highlighted by yellow colour

Amino acid	Synonymous codon	Carassius gibelio	Anisakis simplex
Ala	GCU	**0.393296602**	**0.353917961**
	GCC	0.253998447	0.225695243
	GCA	0.209719515	0.261663301
	GCG	0.142985534	0.158724757
Cys	UGU	**0.540577184**	0.333555243
	UGC	0.44000534	**0.33634767**
Asp	GAU	**0.54819301**	**0.63734767**
	GAC	0.45180699	0.36265233
Glu	GAA	0.373149806	**0.640676019**
	GAG	**0.626850291**	0.359323981
Phe	UUU	0.302271359	0.245591456
	UUC	**0.697728641**	**0.754408544**
Gly	GGU	0.265267961	**0.394228738**
	GGC	0.225051845	0.351060583
	GGA	**0.300926893**	0.209407282
	GGG	0.208752816	0.045303592
His	CAU	0.475683592	**0.572176117**
	CAC	**0.51460767**	0.427823883
Ile	AUU	0.265762524	0.263753689
	AUC	**0.585068932**	**0.613995728**
	AUA	0.149168835	0.122250777
Lys	AAA	**0.558907864**	**0.563443204**
	AAG	0.441092136	0.436556796
Leu	UUA	0.049411942	0.096529612
	UUG	0.089229806	**0.297765437**
	CUU	0.134835437	0.192543786
	CUC	0.213534854	0.222009903
	CUA	0.046709417	0.050579806
	CUG	**0.466278252**	0.140572913
Asn	AAU	0.417444563	0.415921942
	AAC	**0.582555437**	**0.584078058**
Pro	CCU	0.237304563	0.14579699
	CCC	0.23103233	0.110068738
	CCA	**0.364309126**	**0.524651068**
	CCG	0.167354369	0.209774175
Gln	CAA	0.291447767	**0.636111165**
	CAG	**0.708552233**	0.363888835
Arg	CGU	0.115636505	**0.206030874**
	CGC	0.130375631	0.070011942
	CGA	0.097428252	0.176361165

Table continued

that affect the codon usage pattern in addition to base composition. In both the organisms, some genes were positioned in close proximity the standard line that reflects the effect of mutation and compositional constraints. However, there are genes, which were deviated well below and above the line indicating the role of natural selection on codon usage pattern (Liu *et al.,* 2012) (Fig 1 A (a), (b)). Consequently, the codon usage pattern of genes in both host and parasite might be shaped by the combined effects of directional mutation and neutral selection. Neutrality plot depicts the correlation between GC12 and GC3, which further reflects the role of evolutionary forces

(mutation/selection) on codon usage. It has been observed that the slope of regression line was higher than 1 for both the organisms (*C. gibelio* =1.74 and A. simplex = 1. 95), indicating the dominant role of mutation on codon usage pattern (Fig. 2 a, b).

Also, considering the broad GC range in both the organisms the effect of selection pressure could not be ruled out. Altogether, the codon usage pattern of *C. gibelio* and *A. simplex* are shaped both by mutational pressure and translational selection (Yu *et al.,* 2015). The neutrality plot analysis of *C. gibelio* and *A. simplex* is shown in Fig 1 B (a), (b), respectively. Parity plot

Table 2: RSCU values analysis of *Carassius gibelio* and *Anisakis simplex*. The shared codons between *C. gibelio* and *A. simplex* are highlighted by yellow colour.

Amino acid	Synonymous codon	*Carassius gibelio*	*Anisakis simplex*
	CGG	0.08489	0.048478641
	AGA	**0.413047767**	0.134439903
	AGG	0.158623689	0.05399699
Ser	UCU	**0.250840777**	0.070650971
	UCC	0.170892039	0.066128544
	UCA	0.199625728	**0.316080388**
	UCG	0.058041165	0.228222233
	AGU	0.137715922	0.13045835
	AGC	0.18288466	0.188459126
Thr	ACU	0.293330388	0.252799126
	ACC	0.257508932	0.242111456
	ACA	**0.324981359**	**0.336066699**
	ACG	0.124179903	0.169023301
Val	GUU	0.215319806	0.302327087
	GUC	0.309241456	**0.350657573**
	GUA	0.067372136	0.147335243
	GUG	**0.408067087**	0.199680485
Tyr	UAU	0.435172816	0.382321262
	UAC	**0.564827184**	**0.569135049**

Table 3: Correlation between COA Axis 1 and CUB parameters

Axis1	AT3	GC3	ENC
Carassius gibelio	r=-0.312	r = -0.312	r = 0.068
	p< 0.299	p=0.299	p= 0.592
Anisakis simplex	r=-0.065	r= -0.065	r= -0.125
	p= 0.608	p<0.068	p< 0.068

Table 4: Correlation analysis between CAI and several other CUB indices

CAI	A3	T3	AT3	GC12	ENC	GC3	GC1	GC2	Fop
Carassius gibelio	r=-0.418	r = -0.012	r = 0.015	r = 0.128	r = -0.331	r = -0.015	r = 0.149	r= 0.046	r= 0.810
	p< 0.005	p=0.904	p= 0.882	p=0.197	p< 0.005	p= 0.882	p= 0.133	p= 0.641	p< 0.005
Anisakis simplex	r=-0.272	r=-0.370	r= 0.410	r= 0.638	r= -0.467	r= -0.410	r=0.598	r=0.609	r=0.826
	p= 0.005	p<0.005	p< 0.005	p<0.005	p<0.005	p<0.005	p< 0.005	p<0.005	p<0.005

is particularly useful in determining the role of evolutionary forces influencing codon usage bias. In this analysis, it was observed that AT and GC were not proportionally used and such codon usage disproportion between A + T and G + C at third codon position indicates that along with the mutation, selection and other factors also determine the codon usage patterns in *C. gibelio* and *A. simplex* (Shang *et al.,* 2011). PR2 analysis is shown in Fig 1 C (a), (b) for *C. gibelio* and *A. simplex*.

Determination of amino acid composition is an important aspect to be considered, as it is affected by the genomic GC content and in turn affects the codon usage pattern of the organisms (Lightfield *et al.,* 2011, Khrustalev and Barkovsky

2012). In the present study, the most copious amino acid in *C. gibelio* was found to be Ser and Leu followed by Ala, Thr and Val whereas in *A. simplex,* Ala was found to have maximum usage frequency followed by Leu and Lys. It was found that 6-fold and 4-fold amino acids occurred in abundance in both the host and the pathogen. Cys, His and Tyr were among the lowest frequency amino acids in both *C. gibelio* and *A. simplex*. It has been reported that both serine and alanine amino acids are the constituents of SAP (Serine, Alanine and Proline) protein gene family. SAPs are known to play an important role in parasite internalization in host environment by triggering the host cell's Ca^{2+} response (Renata *et al.,* 2006). In view of this, it could be hypothesized that the abundance of serine and alanine in host and parasite might play

Table 5: Correlation analysis between ENC and several other CUB indices

ENC	A3	T3	AT3	GC12	CIA	GC3	GC1	GC2	Fop
Carassius gibelio	r=-0.090 p=0.363	r = -0.267 p=0.006	r= 0.442 p<0.005	r = 0.127 p=0.202	r = -0.331 p< 0.005	r = -0.442 p<0.005	r = 0.117 p= 0.241	r= 0.111 p= 0.266	r= -0.222 p=0.023
Anisakis simplex	r=0.278 p<0.005	r=0.293 p<0.005	r= 0.181 p=0.067	r= -0.684 p<0.005	r= -0.467 p<0.005	r= -0.181 p=0.067	r= -0.608 p< 0.005	r= -0.684 p<0.005	r= -0.218 p=0.027

Table 6: Top 20 preferred codon pairs. Preferred codons followed by identical codons were highlighted in yellow-colored background. All the codon pairs comprised of preferred codons were highlighted in red colored font

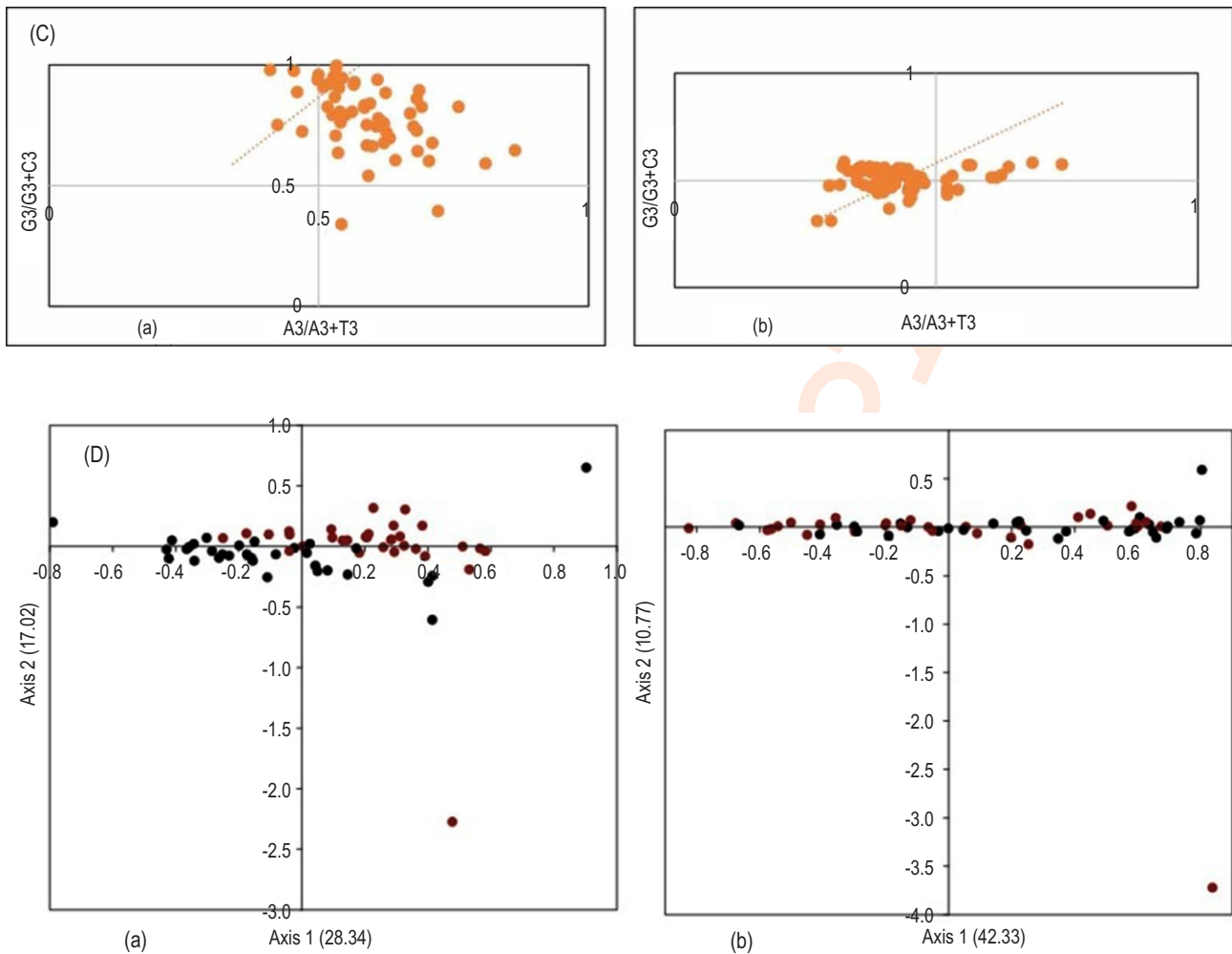| Anisakis simplex	*Carassius gibelio*	Anisakis simplex	*Carassius gibelio*
GAA-GAA	ACU-ACA	AAA-GAG	AAC-AUC
GCU-GAA	GAA-GAU	GAG-AAG	ACA-GCU
GAU-GAA	ACA-GCA	GAA-GAU	AAA-CCA
GAU-GAU	GAU-GAU	AAA-GAA	GAG-GAG
GAU-GCU	GUC-AAA	GUU-GAU	CUG-AUC
GCU-GAU	AAA-GAU	GGU-GAU	ACA-ACA
AAG-AAA	GAU-AAA	GAG-GAU	UCU-GUG
CAA-CAA	CAG-AAA	GCA-CAA	GCU-CCA
CCA-CCA	UCU-CUG	GAA-GAG	GAC-UCU
GAU-GCA	GAG-CAG	GAA-CUG	AAG-AAG

(A)

(a) (b)

(B)

(a) (b)
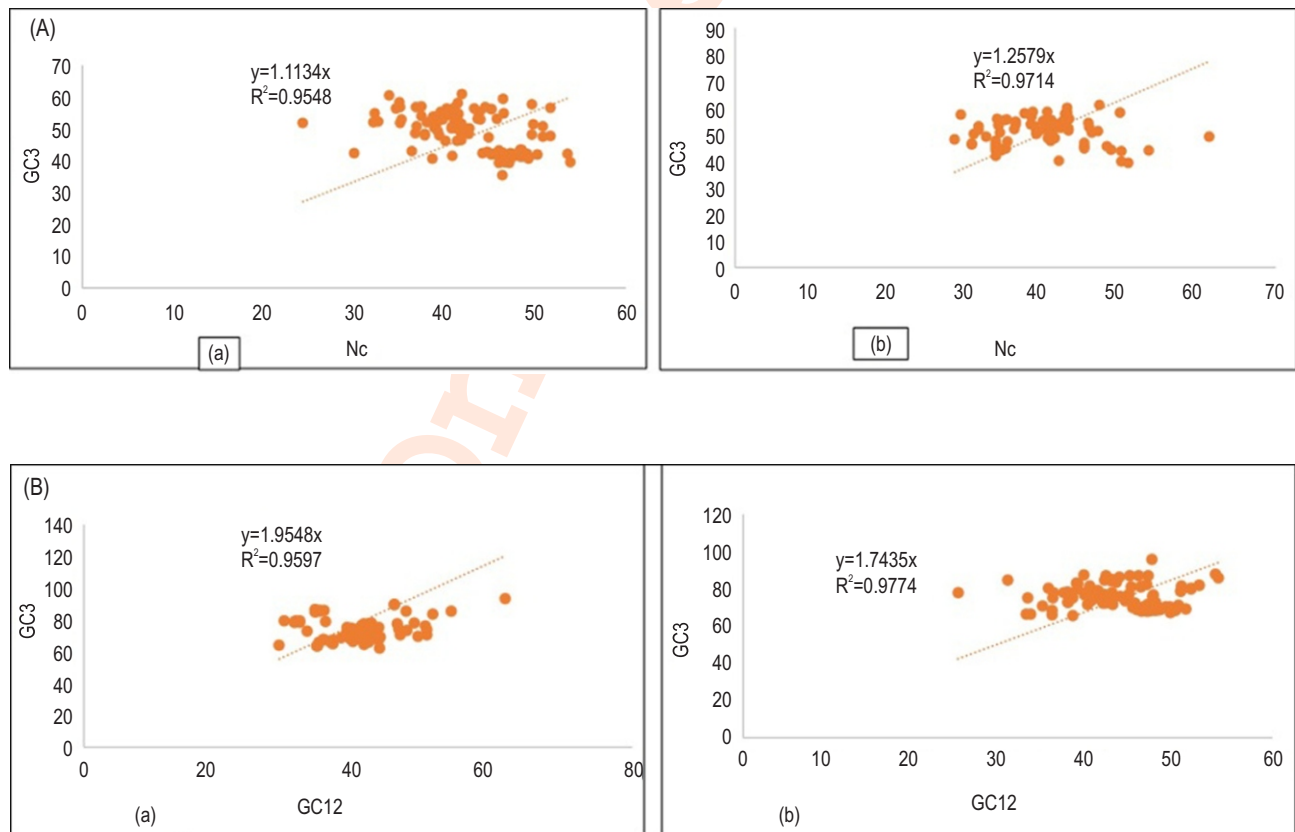
Fig. 1: Different plots showing influence of evolutionary forces on codon usage pattern of *Carassius gibelio* (a) and *Anisakis simplex* (b) (A) ENC plot (B) Neutrality plot (C) Parity plot (D) COA analysis.

role in their interaction and thus would be involved or helpful in establishing parasitism. The amino acid usage frequency is shown in Fig 2. The correspondence analysis (COA) was performed to understand what other factors affects the codon usage pattern in *C. gibelio* and *A. simplex*. This multivariate statistical method reviews variation of RSCU values within the genome (Greenacre, 1984). The correspondence analysis shows the distribution of genes and their corresponding codons, revealing potential influence on CUB (Romero *et al.,* 2000).

Correspondence analysis of *C. gibelio* and *A. simplex* was carried out based on the RSCU values. Among 59 orthogonal axes that represented 59 codons, axis 1 followed by axis 2 accounted as the major contributors of total variation in codon usage pattern in *C. gibelio* and *A. simplex,* and were thus selected for further analysis. Positioning of codons along both the axis reflects that evolutionary forces experienced by each codon was

different. A few codons were clustered around the origin of the axis indicated similar CUB operative in both the species. In *C. gibelio*, most of the codons were G/C ending (indicated by brown dots) and were positioned at the positive side of the plot whereas A/U ending codons were found at the negative side of the plot (indicated by black dots). In *A. simplex*, G/C ending codons (brown dots) were dominant at the positive side of the plot similar to *C. gibelio*. Conversely, A/U ending codons (black dots) were spread almost equally on the negative as well as the positive sides of the plot. Correlation study was conducted between axis 1 and other CUB affecting parameters like, ENC, GC3 and AT3 content (Table 3). Non-significant correlation was observed between axis 1 and other parameters in *C. gibelio* and *A. simplex*. The position of each codon on the plane defined by the first two axes for *C. gibelio* and *A. simplex* is displayed in Fig. 1 D (a, b) respectively. Overall the COA analysis elucidates the combined role of mutational pressure and selection in inducing total
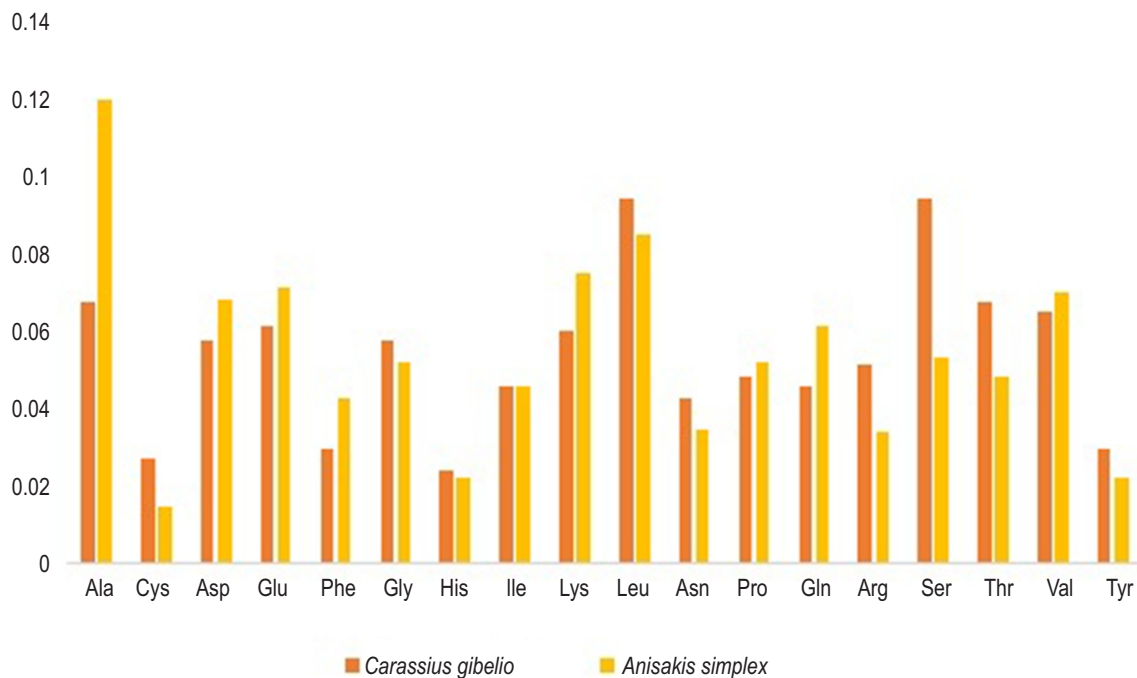
Fig. 2: Amino acid frequency usage pattern in *Carassius gibelio* and *Anisakis simplex*.

variation in codon usage pattern in *C. gibelio* and *A. simplex* and first axis being the major contributor to overall variation in both species (Musto *et al.,* 1998; Jia *et al.,* 2015). The correlation analysis was performed between different codon usage bias (CUB) indices (used in the present study) to gain insight into the factors influencing codon usage pattern in *C. gibelio* and *A. simplex* (Table 4, 5). In *C. gibelio*, CAI was positively correlated with AT3, GC12, GC1, GC2 and Fop, while negatively correlated with A3, T3, CAI, GC3 and FOP. Among all the Aindices, FOP showed the highest association with CAI. Similar correlation pattern was also observed in *A. simplex* between CAI and other CUB indices. The correlation of ENC with AT3, GC12, GC1 and GC2 was in the positive direction in *C. gibelio* whereas in *A. simplex* only three compositional indices, *i.e.,* A3, T3 and AT3 showed positive correlation with ENC. Interestingly, AT3 in *C. gibelio* and GC12 in *A. simplex* showed the most significant correlation with CUB. Analysis related to codon-codon pairing was done on top 20 high-frequency codon pairs out of the 64×64 codon pairs.

Similar trend of codon pairing has been observed in both host and parasite. Most of the codon pairs consisted of preferred codons (marked by red color) (Table 6). Moreover, the tendency towards the usage of identical synonymous codon pairs (highlighted with yellow colour) was also observed in both the species. Similar findings have been documented, where organisms have shown their inclination towards preferred and similar codon pairing (Roy and Staden, 2019).The present study also highlighted that codons of certain amino acids were not represented among the top 20 high-frequency codon pairs. In *C. gibelio*, the codons of Cys, Phe, Gly, His, Arg amino acids were absent, while in *A. simplex*, the codons of amino acids Cys, Phe, His, Ile, Leu, Asn, Arg, Thr, Val and Tyr were missing. The absence of some selected codon pairs in both the fish and nematode genomes could be attributed to the expense incurred in the biosynthesis of corresponding amino acids (Raiford *et al.,* 2008; Williford and Demuth, 2012).

The present study highlights the codon usage pattern operative in fish *C. gibelio* and its nematode parasite *A. simplex*. Through various computational tools similar pattern of codon usage bias was observed in the coding sequences of *C. gibelio* and *A. simplex* (AC-ending codons). Based on these results, we can hypothesize that codon usage bias plays a major role in the successful colonization of parasite inside the host. In addition, inclusive analysis using different plots (ENC, parity, neutrality) has shown the influence of both the evolutionary forces *i.e.,* mutational and translational selection on codon usage pattern. The results of the present study would lay a foundation for future research on other parasites associated with carps and other fishes belonging to Cyprinidae family. This study gives us an understanding into the particular genetic features common in the genomes of *A. simplex* and *C. gibelio*.

research facilities. We are also grateful to Negenome Bio Solutions Pvt. Ltd. for their constant support to carry out research work.

Add-on Information

References

Akashi, H.: Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics*, **136**, 927-935 (1994).

Angellotti, M.C., S.B. Bhuiyan, G. Chen and X.F. Wan: CodonO: codon usage bias analysis within and across genomes. *Nucleic Acids Res.,* **35** (Suppl_2), W132-W136 (2007).

Biswas, K.K., S. Palchoudhury, P. Chakraborty, U.K. Bhattacharyya, D.K. Ghosh, P. Debnath and R.F. Lee: Codon usage bias analysis of Citrus tristeza virus: Higher codon adaptation to *Citrus reticulata* host. *Viruses*, **11**, 331 (2019).

Bullini, L.,P. Arduino, R. Cianchi, G. Nascetti, S. D'amelio, S. Mattiucci and J. Smith: Genetic and ecological research on anisakid endo-parasites of fish and marine mammals in the Antarctic and Arctic-boreal regions. In: Antarctic communities: Species, structure and survival. (Eds.: B. Battaglia, J. Valencia and D.W.H. Walton). Cambridge University Press, Cambridge, pp. 118-134 (1997).

Butt, A.M., I. Nasrullah and Y. Tong: Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. *PLOS OONE,* **93**, e90-905 (2014).

Crivelli, A.J.: Are fish introductions a threat to endemic freshwater fishes in the Northern Mediterranean region? *Biol. Conserv.,* **72**, 311–319 (1995).

Cutter, A.D., J.D. Wasmuth and M.L. Blaxter: The evolution of biased codon and amino acid usage in nematode genomes. *Mol. Biol. Evol.,* **23**, 2303-2315 (2006).

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).

Duret, L.: Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.,* **12**, 640-649 (2002).

Frumkin, I., M. J. Lajoie, C. J. Gregg, G. Hornung, G. M. Church and Y. Pilpel: Codon usage of highly expressed genes affects proteome-wide translation efficiency. Proceedings of the National Academy of Sciences, **115**, E4940-E4949 (2018).

Greenacre, M.J.: Theory and Applications of Correspondence Analysis. Academic Press, London (1984).

Gupta, S. K. and T.C. Ghosh: Gene expressivity is the main factor in dictating the codon usage variation among the genes in *Pseudomonas aeruginosa. Gene,* **273**, 63-70 (2001).

Gupta, S.K., T.K. Bhattacharyya and T.C. Ghosh: Synonymous codon usage in *Lactococcus lactis*: Mutational bias versus translational selection. *J. Biomol. Struct. Dyn.,* **21**, 527-535 (2004).

Hershberg, R. and D.A. Petrov: General rules for optimal codon choice. *PLoS Genet.,* **5**, e1000556 (2009).

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

Jia, X., S. Liu, H. Zheng, B. Li, Q. Qi,, L. Wei and J. Sun: Non-uniqueness of factors constraint on the codon usage in *Bombyx mori. BMC Genomics,* **16**, 1-12 (2015).

Kahali, B., S. Basak and T.C. Ghosh: Delving deeper into the unexpected correlation between gene expressivity and codon usage bias of *Escherichia coli* genome. *J. Biomol. Struct. Dyn.,* **25**, 655-661 (2008).

Kawabe, A. and N.T. Miyashita: Patterns of codon usage bias in three dicot and four monocot plant species. *Genes Genet. Syst.,* **78**, 343-352 (2003).

Khrustalev, V.V. and E.V. Barkovsky: Stabilization of secondary structure elements by specific combinations of hydrophilic and hydrophobic amino acid residues is more important for proteins encoded by GC-poor genes. *Biochimie,* **94**, 2706-2715 (2012).

Kiewitz, C. and B. Tümmler: Sequence diversity of *Pseudomonas aeruginosa*: Impact on population structure and genome evolution. *J. Bacteriol.,* **182**, 3125-3135 (2000).

Kottelat M. and J. Freyhof: Handbook of European Freshwater Fishes. Kottelat, Cornol, Switzerland and Freyhof, Berlin, Germany, 646 Pages (2007).

Lightfield, J., N.R. Fram and B. Ely: Across bacterial phyla, distantly-related genomes with similar genomic GC content have similar patterns of amino acid usage. *PloS ONE,* **6**, e17677 (2011).

Liu, H., R. He, H. Zhang, Y. Huang, M. Tian and J. Zhang: Analysis of synonymous codon usage in Zea mays. *Mol. Biol. Rep.,* **37**, 677 (2010).

Liu, H., Y. Huang, X. Du, Z. Chen, X. Zeng, Y. Chen and H. Zhang: Patterns of synonymous codon usage bias in the model grass *Brachypodium distachyon. Genet. Mol. Res.,* **11**, 4695-4706 (2012).

Ma, Y.P., Z.X. Liu, L. Hao, J.Y Ma, Z.L. Liang, Y.G. Li and H. Ke: Analysing codon usage bias of cyprinid herpesvirus3 and adaptation of this virus to the hosts. *J. Fish Dis.,* **38**, 665- 673 (2015).

Mitra, S., S. K. Ray and R. Banerjee: Synonymous codons influencing gene expression in organisms. *Rese. Rep. Biochem.*, **6**, 57-65 (2016).

Moravec, F.: Nematodes of freshwater fishes of the Neotropical Region. Academia, Publishing House of the Academy of Sciences of the Czech Republic, 464 Pages (1998).

Mougel, F., C. Manichanh, G.D. N'guyen and M. Termier: Genomic choice of codons in 16 microbial species. *J. Biomol. Struct. Dyn.,* **22**, 315-329 (2004).

Musto, H., H. Romero and H. Rodríguez-Maseda: Heterogeneity in codon usage in the flatworm *Schistosoma mansoni. J. Mol. Evol.,* **46**, 159-167 (1998).

Nasrullah, I., A.M. Butt, S. Tahir, M. Idrees and Y. Tong: Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. *BMC Evol.*

Biol., **15**, 1-15 (2015).

Pop, C., S. Rouskin, N.T. Ingolia, L. Han, E.M. Phizicky, J.S. Weissman, and D. Koller: Causal signals between codon bias, mRNA structure and the efficiency of translation and elongation. *Mol. Syst. Biol.,* **10**, 770 (2014).

Quax, T.E., N.J. Claassens, D. Söll and J. van der Oost: Codon bias as a means to fine-tune gene expression. *Mol. Cell.,* **59**, 149-161 (2015).

Raiford, D.W., E.M. Heizer, R.V. Miller, H. Akashi, M.L. Raymer and D.E. Krane: Do amino acid biosynthetic costs constrain protein evolution in *Saccharomyces cerevisiae*? *J. Mol. Evol.,* **67**, 621-630 (2008).

Renata C.P. Baida, Ma´rcia R.M. Santos, M.S. Carmo, N. Yoshida, D. Ferreira, A.T. Ferreira, N.M. El Sayed, B. Andersson and J.F. da Silveira: Molecular characterization of Serine-, Alanine-, and Proline-Rich proteins of *Trypanosoma cruzi* and their possible role in host cell infection. *Infect. Immun.,* **74**, 1537–1546 (2006).

Romero, H., A. Zavala and H. Musto: Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. *Nucleic Acids Res.,* **28**, 2084-2090 (2000).

Roy, A. and J. van Staden: Comprehensive profiling of codon usage signatures and codon context variations in the genus *Ustilago. World J. Microbiol. Biotechnol.,* **35**, 118 (2019).

Sahoo, S., S.S. Das and R. Rakshit: Codon usage pattern and predicted gene expression in *Arabidopsis thaliana. Gene X.,* **2**, 100012 (2019).

Sahu, K., S.K. Gupta, S. Sau and T.C. Ghosh: Comparative analysis of the base composition and codon usages in fourteen mycobacteriophage genomes. *J. Biomol. Struct. Dyn.,* **23**, 63-71 (2005).

Shang, M., F. Liu, J. Hua and K. Wang: Analysis on codon usage of chloroplast genome of *Gossypium hirsutum. Zhongguo Nong Ye Ke Xue,* **44**, 245-253 (2011).

Sharma J., S. Chakraborty and A. Uddin: Comparative analysis of codon usage bias between two Lepidopteran insect species: *Bombyx mandarina* and *Ostrinia furnacalis. Int. J. Sci. Res.,* **3**, 47–50 (2014).

Sharp, P.M. and W.H. Li: An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.,* **24**, 28-38 (1986).

Sueoka, N.: Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci. USA.,* **85**, 2653-2657 (1988).

Sueoka, N.: Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. *J. Mol. Evol.,* **40**, 318-325 (1995).

Tuller, T., Y.Y. Waldman, M. Kupiec and E. Ruppin: Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci.,* **107**, 3645-3650 (2010).

Williford, A. and J.P. Demuth: Gene expression levels are correlated with synonymous codon usage, amino acid composition and gene architecture in the red flour beetle, *Tribolium castaneum. Mol. Biol. Evol.,* **29**, 3755-3766 (2012).

Wong, E.H., D.K., Smith, R. Rabadan, M. Peiris and L.L. Poon: Codon usage bias and the evolution of influenza A viruses. Codon usage biases of influenza virus. *BMC Evol. Biol.,* **10**, 1-14 (2010).

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

Zhao, K.N., W.J. Liu and I.H. Frazer: Codon usage bias and A+ T content variation in human papillomavirus genomes. *Virus Res.,* **98**, 95-104 (2003).