BIOINFORMATIC ANALYSIS OF MITOCHONDRIAL GENOME OF HUMPED FEATHERBACK CHITAL FISH, *Chitala chitala* (HAMILTON, 1822)

MD. HASIN IRTIZA

DEPARTMENT OF BIOTECHNOLOGY

SHER-E-BANGLA AGRICULTURAL UNIVERSITY

DHAKA-1207

December, 2021

BIOINFORMATIC ANALYSIS OF MITOCHONDRIAL GENOME OF HUMPED FEATHERBACK CHITAL FISH, *Chitala chitala* (HAMILTON, 1822)

BY

MD. HASIN IRTIZA

REGISTRATION NO.: 19-10313
Mobile no. 01715597556
E-mail:hasinirtiza1.hi@gmail.com

A Thesis
Submitted to the Faculty of Agriculture,
Sher-e-Bangla Agricultural University, Dhaka,
in partial fulfilment of the requirements
for the degree of

MASTER OF SCIENCE (MS)
IN
BIOTECHNOLOGY

SEMESTER: JULY-DECEMBER, 2021

Approved by:

(Dr. Mohammad Nazrul Islam) Associate Professor Department of Biotechnology Faculty of Agriculture Sher-e-Bangla Agricultural University, Dhaka-1207 **Supervisor**	**(Shirin Sultana)** Senior Scientific Officer Fisheries Biotechnology Division National Institute of Biotechnology Savar, Dhaka-1349 **Co-supervisor**

(Dr. Mohammad Nazrul Islam)
Chairman and Associate Professor
Examination Committee
Department of Biotechnology
Sher-e-Bangla Agricultural University, Dhaka-1207

Department of Biotechnology

Sher-e-Bangla Agricultural University
Sher-e-Bangla Nagar
Dhaka-1207

CERTIFICATE

This is to certify that the thesis entitled, **"BIOINFORMATIC ANALYSIS OF MITOCHONDRIAL GENOME OF HUMPED FEATHERBACK CHITAL FISH, *Chitala chitala* (HAMILTON, 1822)"** submitted to the Faculty of Agriculture, Sher-e-Bangla Agricultural University, Dhaka-1207 in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE (MS) IN BIOTECHNOLOGY**, embodies the result of a piece of bona fide research work carried out by **MD. HASIN IRTIZA,** Registration No.**19-10313** under my supervision and guidance. No part of the thesis has been submitted for any other degree or diploma.

I further certify that such help or source of information, as has been availed of during the course of this investigation has duly been acknowledged.

Dated: December, 2021
Dhaka, Bangladesh

Dr. Mohammad Nazrul Islam
Associate Professor
Department of Biotechnology
Sher-e-Bangla Agricultural University
Dhaka-1207
Supervisor

DEDICATED TO
MY
BELOVED PARENTS

LIST OF ABBREVIATIONS

ABBREVIATION	ELABORATION
Mitogenome	Mitochondrial genome
DNA	Deoxyribonucleic acid
bp	Base pair
A	Adenine
T	Thymine
C	Cytosine
G	Guanine
tRNA genes	Transfer RNA genes
PCGs	Protein-coding genes
rRNA genes	Ribosomal RNA genes
D-loop	Displacement loop
NCBI	National Center for Biotechnology Information
BLAST	Basic Local Alignment Search Tool
GenBank	GenBank sequence database
NIB	National Institute of Biotechnology
CAI	Codon adaptation index
ENc	Effective number of codon
%GC	Percentages of GC
% GC1	Percentages of GC content at the codon position 1^{st}
% GC2	Percentages of GC content at the codon position 2^{nd}
% GC3	Percentages of GC content at the codon position 3^{rd}
RSCU	Relative synonymous codon usage
Ka	Non-synonymous substitution
Ks	Synonymous substitution
ω	Non-synonymous substitution/Synonymous substitution ratio
SNAP	Synonymous non-synonymous Analysis Program
OL	Origin of light strand
IGS	Intergenic spacers
tRNA-Asn	Transfer RNA-Asparagine
tRNA-Cys	Transfer RNA-Cysteine
kcal/mole	Kilo calorie per mole
$\delta\delta G$	The free energy increment
ML method	Maximum Likelihood method

LIST OF ABBREVIATIONS (CONTINUED)

ABBREVIATION	ELABORATION
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
Cys	Cysteine
Gln	Glutamine
Glu	Glutamic acid
Gly	Glysine
His	Histidine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine
Met	Metheonine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine

ACKNOWLEDGEMENTS

In the name of Allah, the most Gracious and the most Merciful.

All the praises are to the supreme being, creator and ruler of the universe Allah, whose mercy keeps me alive and enables me to pursue my education in the Department of Biotechnology, Sher-e-Bangla Agricultural University and to complete my thesis paper for the Degree of Master of Science in Biotechnology.

Firstly, I would like to express my utmost indebtedness and gratitude to my research supervisor **Dr. Mohammad Nazrul Islam**, Associate Professor and Chairman Department of Biotechnology, Sher-e-Bangla Agricultural University, for his careful supervision, sincere guidance, constructive criticism throughout the course and preparing this thesis paper as well.

I also feel pleasure to express my sincere respect and cordial thanks to my research co-supervisor, **Shirin Sultana**, Senior Scientific Officer, Fisheries Biotechnology Division, National Institute of Biotechnology, for her utmost co-operation for completing the thesis.

I would also like to thank all the teachers and staffs of the department due to providence of all necessary facilities.

Finally, I would like to thank my parents, who had been sacrificed their happiness for me and rendered financial support and encouragement throughout my academic career. This research work would not have been possible without them. I will be indebted to them forever. I will try my best to keep them happy in the future. Lastly, I am delighted to avail myself to express my sincerest thanks to all the classmates.

Dated: December, 2021 The Author

LIST OF CONTENTS

LIST OF CONTENTS (CONTINUED)

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

BIOINFORMATIC ANALYSIS OF MITOCHONDRIAL GENOME OF HUMPED FEATHERBACK CHITAL FISH, *Chitala chitala* (HAMILTON, 1822)

ABSTRACT

The present study was about the bioinformatic analysis of mitochondrial genome of an endangered featherback fish, *Chitala chitala.* Although *Chitala chitala* plays significant role in inland fishery, but its' evolution has received little attention. In the current study, the mitogenome of *Chitala chitala* was retrieved and investigated. Some of the bioinformatics tools were applied to measure important gene parameters including AT/GC-skewness, codon adaptation index (CAI), the effective number of codons (ENc) and GC percentages of each protein coding gene. The majority of AT-skew and GC-skew values of all protein coding genes of the species were negative, and the amplitude of the GC-skew was larger than the AT-skew. On the basis of comparative selective strength analysis with the PCGs of two related mitogenomes, I discovered that most of the PCGs retained the Ka/Ks ratios less than one implying that they evolved under strong negative or purifying selection. Moreover, I analyzed the codon frequency and relative synonymous codon usage (RSCU). Then, I identified a total of 22 RSCU values (>1) and revealed 6 codons as "over-presented" that implied for codon usage bias to engage in highly expressed genes for efficient protein synthesis via translational selection. The phylogenetic tree and time tree were constructed by Maximum Likelihood (ML) method, providing further supplement to the evolution of the fish. Regarding on the facts of the presence of codon usage biasness rolling in translational selection and the signs of purifying selection identified in PCGs indicate obvious conservation of this endangered fish species.

CHAPTER I
INTRODUCTION

Chitala chitala (Hamilton, 1822) is an evolutionarily significant freshwater primitive fish, which is locally familiar as Chital. It belongs to the order of Osteoglossiformes and the family of Notopteridae. It is distributed in the Indian subcontinent along with Bangladesh, India, Pakistan, Nepal and Myanmar. In nature, the dwelling of the fish comprise the rivers, beels, canals, reservoirs, ponds etc. It is called as clown knife fish, giant featherback also (Chonder, 1999; Mitra et al., 2018). High market value has been achieved, not only as food fish but also for ornamental mercantile (Mandal et al., 2012). The population in nature of *C. chitala* has been decreased because of over catching of this fish in its plenitude and is presently enlisted as the category of endangered species (IUCN Bangladesh, 2015).

The family Notopteridae contains 10 species of osteoglossiform (bony-tongued) fishes, commonly known as featherbacks and knifefishes. These fishes live in freshwater or brackish environments in Africa and South and Southeast Asia. Featherbacks have slender, elongated, bodies, giving them a knife-like appearance. The caudal fin is small and fused with the anal fin, which runs most of the length of the body. Where present, the dorsal fin is small and narrow, giving rise to the common name of "featherback". The fish swims by holding its body rigid and rippling the anal fin to propel itself forward or backwards (Greenwood et al., 1998). Notopterids have specialized swim bladders. The organ extends throughout the body and even into the fins in some cases. Although the swim bladder is not highly vascularised, it can absorb oxygen from air and also functions to produce sound by squeezing air through a narrow passage into the pharynx (Greenwood et al., 1998).

Chitala chitala (Hamilton, 1822), the Clown Knifefish of family Notopteridae, pertains to one of the primitive orders, Osteoglossiformes. The fossil records of order Osteoglossiformes have

been retrieved from the deposits belonging to late Jurassic or early Cretaceous period indicating an ancestral teleost lineage (Hilton et al., 2018). *C. chitala* is commonly known as humped featherback due to the long anal fin which is confluent with caudal fin gives an appearance of feather (Chonder, 1999).

The fish has an elongated body. The head and the body are strongly compressed laterally. Dorsal profile is highly convex. Scales are very minute and short dorsal fin. Anal fin is long and confluent with caudal fin. Pectoral fins are reduced. Dorsal portion is coppery green coloured and silvery at sides and below. 15 silvery bars present on each side of dorsal ridge. 5-9 small black spots close to the end of the region of caudal fin. Lateral line is complete. Maximum length reported 120 cm (Day, 1878). In Bangladesh, 103 cm (10 kg) was recorded from Gacher Dahar Beel of Sylhet district (Rahman, 1989).

The fish is carnivorous and predator (Rahman,1989). It feeds on aquatic insects, molluscs, shrimps and small fishes and takes insects and tender roots of aquatic plants during its earlier stage of life (Bhuiyan, 1964).

It breeds between June and July by building nest and eggs are 3-4.5 mm in diameter (Rahman, 1989). Eggs are 3-4.5 mm in diameter and young receive parental care (Bhuiyan, 1964).

The fish is used as food fish in Bangladesh. Always this fish is marketed in fresh and sometimes in live condition. Flesh is of good flavour but full of small bones (Talwar et al., 1991).

The importance of this species is both for conservation and for aquacultural purposes as well (Mandal et al. 2011). Some initial genetic studies had been performed in the past. The evolution of these might have connection to the biogeographical perspectives and these species evolved and did not spread to saline waters (Banarescu 1990; Lundberg 1993). With the view to providing platform for its conservation and management, the evolutionary history should be documented (Desmet et al. 2002; Moritz 2002).

There is usage of mitochondrial DNA, a necessary resource as molecular marker for the exploration of the genetics of population, phylogenetic relationships, genetic barcoding and the histories of evolution.

There is high mutation rate in mitochondrial genomes, because of the free radicals which are propagated by the respiration inside the mitochondria and remain very nearly to the mitochondrial DNA. Hence, mitochondrial genes undergo much higher and faster evolution rates than nuclear genes. Mutation pressure generates variation in the frequencies of the nucleotide bases at both synonymous and non-synonymous sites in the several species' mitogenomes (Jia et al., 2008; Mazumder et al., 2020). The mutation rates are not equal because of the asymmetric type of replication of two strands of mitochondrial genome, which means that the frequencies of G and C are not same, and the frequencies of A and U are also different (Faith et al., 2003).

In spite of the fact that, mitochondrial gene order and composition are highly conserved, nucleotide sequences show swift divergence (Billington et al.,1991). The renowned population geneticist, Motoo Kimura postulated the molecular evolution theory which was neutral. In the theory, it was asserted many occurrences where a genetic change spreads across a population due to neutral mutations caused by genetic drift (Kimura et al.,1968). However, the change is not necessarily emerged in phenotypic change (Kimura et al.,1968). When the condition of natural selection takes place, the selective removal of deleterious alleles occurs through negative selection or purifying selection (Loewe et al., 2010). Primarily deeply recessive alleles (may be deleterious or beneficial) purging by genetic drift can be removed on the contrary, natural selection can remove any type of deleterious alleles (Glémin et al., 2003). In small population, fixation of any selection pressures takes place instantaneously in fewer generations than that in large random mating populations. In case of strong purifying or negative selection on a locus, the purification of deleterious variants sequels to the incidental removal of linked

variation. This remarkable purging of non-deleterious alleles adjacent to deleterious ones is called background selection which may produce decreased level of variation surrounding the locus under selection (Charlesworth et al., 1993).

As a result of the degeneracy of genetic code, the functional expression of a protein in an organism is determined by preference of codon usage. The forms of codon usage alter in different protein-coding genes of a mitogenome (Mazumder et al., 2015). It was the pressure of mutation which was prevalent in determining the usage of codon patterns in mitochondrial genes that can be affected by translational selection (Jia et al., 2008). Other factors including compositional constraint of nucleotides, interaction between codon and anticodon, amino acids' conservation, etc. may impact the frequency of codon, the biasness of codon usage in the genome of an organism (Uddin et al., 2016;Sahoo et al., 2019).

Now-a-days mitochondrial genomes (mitogenome) have become a potential molecule marker to enquire into phylogenetic analysis and taxonomic diagnosis in ichthyological researches due to the simple genetic structure, inheritance maternally, rapid rate of evolution, easy detection and exalted specificity.

For fish species identification, the mitochondrial species-specific DNA fragments, such as ribosomal RNA (12S and 16S), COI and CYTB are used generally (Waldbieser et al., 2003). However, individual mitochondrial gene fragments have displayed a faulty performance in inquiring into the phylogenetic relationship among divergent species and individualizing particular congeneric species. On account of the slow evolution and insufficient nucleotide mutation rate, the COI which is viewed as a 'DNA barcode' cannot identify some closely related fishes. Moreover, either the non-coding region (control region) or CYTB shows weak abilities for the reason of solving relationship in a rapid radiation. Therefore, more insights and better resolution than single mitochondrial sequences in taxonomic level can be provided by the complete mitochondrial genome with greater sequences data.

The shape of mitogenome is usually circular in vertebrates. It is generally comprised of transfer RNA (tRNA) genes, protein-coding genes (PCGs), ribosomal RNA (rRNA) genes, and one control region (D-loop) with a size of approximately 15–20 kb grossly (Chen et al., 2021). For the reason of being the most enormous group of vertebrates, fish has a mitochondrial genome that is alike in composition and structure to most vertebrates. By and large, the array of mitochondrial genes is extremely compacted and highly conserved, but the information offered by mitogenome is distinctive among different species. Nevertheless, the understanding of fish mitochondrial genomes was not deep enough because most researches on fish mitochondrial genome have just simply described the gene structure of single species without thorough comparisons.

The giant featherback belongs the family of Notopteridae (Kottelat et al., 1997). It possesses great financial value and culture. With regard to its flesh has very tasty and distinguishable initially because of high content of fat (Sunarno et al., 2002), and good content of protein and high vitamin A also.

Hence, the giant featherback species would be endangered. Because there are lack of available information on the differences of population and variation genetically among the populations of giant featherback. In reality, most of the previous genetic studies of giant featherback were on the morphological features' basis (Sunarno et al., 2007; Wibowo et al., 2008). There is absence of any research connected to the giant featherback populations' variation in broad areas like rivers, to an elaborated extent, the sequencing of DNA.

In summary, according to the above accounts and scenarios, as there are serious and alarming threats to the existence of various fish species in Bangladesh including *Chitala chitala* have been endangered due to habitat degradation and overexploitation of natural populations. The *C. chitala* a freshwater featherback fish which was categorized as an endangered species in Bangladesh (IUCN 2015). In the case of present study, base composition and skewness, various

gene parameters, codon frequency and relative synonymous codon usage (RSCU), selection pressure, mutation rate and phylogenetic relationships had been analyzed for the thirteen protein-coding genes of the mitogenome of the endangered featherback fish, *C. chitala*. Therefore, this study would provide important information on codon usage patterns and mutation pressure dominant on the fish species.

Considering the above mentioned importance of mitochondrial genome, the current study was designed to explore the following objectives:

i. To retrieve bioinformatic data of mitochondrial genome of Chital fish.
ii. To analyse the mitogenome of Chital fish to find out codon usage bias, mutation rate and purifying selection.
iii. For analysing phylogenetic relationship of Chital fish with its' relatives.

CHAPTER II
REVIEW OF LITERATURE

A considerable literature is available on the mitochondrial genome analysis of different fishes. An attempt is made to review the available literature those were relevant with the present study in this chapter are presented below.

Huiting et al. (2020) conducted study on the mitogenomes of the three species of *Gerres* (*Gerres filamentosus*, *Gerres erythrourus*, and *Gerres decacanthus*) and which were methodically researched. The sizes of the sequences of mitogenome were 16,673, 16,728, and 16,871 bp for *G. filamentosus*, *G. erythrourus, and G. decacanthus,* gradually. The majority of protein-coding genes (PCGs) were started respectively with the ATG codon and terminated with the TAA codon, and the incomplete termination codon TTA could be found in the three species. The values were negative of AT and GC skews of the thirteen protein-coding genes within the three species to the greater extent. The magnitude of GC skew was greater than the value of AT skew. After the exploration of genetic distance and Ka/Ks proportion pointed out that the 13 PCGs were undergoing negative selection. They showed varied pressures of selection from deep-sea fishes. The reason was because of the dissimilation in their dwelling habitat. The creation of phylogenetic tree was done by molecular method, showing that, the three *Gerres* species were segregated in late Cretaceous and early Paleogene. The geological events that could alter their survival environment might be linked with the evolution of these three species.

Huirong et al. (2018) reported the classification of the sequences of the mitogenome of three croakers which were in the order Perciformes and the family Sciaenidae. The distribution and the gene array of these three croakers were similar and congruent to other vertebrates. As the family Sciaenidae was diversified from Perciformes order and did not pertain to any acquainted order, so it was a distinct branch. The division of the distinct Sciaenidae order was retreating

than the Perciformes family. The analysis provided new knowledge on the phylogenetic analysis of the Sciaenidae family from the Perciformes order. And it simplifies further studies on phylogeny of the series Eupercaria and the evolution as well.

Islam et al. (2020) conducted study on mitogenome sequencing and characterization of *Cirrhinus reba* where the samples were collected from the Khulna division of Bangladesh. The mitogenome length was 16597 bp. It shaped as a circular double stranded DNA molecule which contained 37 genes in total. The structure was identical to the other fishes of Teleostei. After the construction of the phylogenetic tree displaying that it had intimate kinship with other labeonine fishes.

Islam et al. (2022) analyzed comparative selective strength with protein coding genes of the three mitogenomes (*Cirrhinus reba*, *Labeo calbasu*, *Labeo bata*) elucidated that 10 PCGs (excluding ND1, COX2 and ND5 genes exhibited the very small Ka/Ks values (<0.1) indicating that the genomes of *C. reba* evolved through strong negative or purifying selection. They also identified a total of 25 RSCU values greater than one and sorted out 12 codons over-presented which indicated clear codon usage bias involved in lightly expressed genes through translational selection of respective codons.

Dutta et al. (2020) reported, *Chitala chitala* has great significance in respect to aquacultural perspective and preservation of its endangered gene pool. This research revealed upper rate of differentiation which was genetic with signs for taking place of four original genetic stocks. Population related scales suggested that the necessity of assessment for a fine scale structuring by applying multi locus markers especially in Gangetic rivers. Whereas, the rest of rivers Brahmaputra and Gomti showed lack of diversities. And in the Mahanandi river, alternate haplotype was absent. For conservation, these populations were vital. For this, the hatchery

managers had to take compatible steps. Finally, by taking adequate measures and with the help of policy makers, researches and stakeholders, this might be attained.

Anjarsari et al. (2021) analysed partial sequence size of the COI gene (621 bp) of featherback. After doing BLAST analysis, species from Indralaya showed the maximum identicality of 99.28% to *Chitala lopis* from Malaysia, than 98.88% to the same species from Kampar River. Whereas, Musi River species showed the highest similarity (95.19%) with *Chitala chitala* from India. The construction of phylogenetic tree indicated that Chitala created four sub-clades which was unambiguously distant from the *C. chitala* and *C. lopis* species.

Singh et al. (2019) conducted a study on the mitogenome of *Chitala chitala*. The sequencing and mapping of the mitogenome were carried out. The size of the mitogenome was 16375 bp. The proportion of Ka/Ks referred that ten genes had negative selection in the developmental process. The construction of phylogenetic trees were done with twelve PCGs to ensure its taxonomic connectivity with other seven orders including osteoglossiformes. Finally, *C. chitala* demonstrated monophyly relationship with others.

Mandal et al. (2011) described, the genetic construction of *Chitala chitala* population that was impacted by the model of genetic diversity, haplotype networks which definitely pointed out two separate lineages that were mitochondrial and mismatch distribution having impacts on the population. The minimum starting point used for comparison information on genetic variation and the proof of sub-structuring of the population produced from this study would be beneficial for feasible strategies for the conservation and exoneration of this endangered species.

Inoue et al. (2009) conducted a study on the mitogenome from 10 osteoglossomorph fishes to determine phylogenetic kinships applying Bayesian and ML methods and the dates of divergence of the family Notopteridae by using a partitioned Bayesian method. 6 species from Notopteridae descent and from the remainder of the family of osteoglossomorph 7 species had

been used by them. As outgroups, 14 more derived teleosts, 9 basal actinopterygians, 2 coelacanths, and 1 shark were used. The phylogenetic exploration suggested that, the African and Asian notopterids formed a sister branch to each other and that these notopterids were a sister to a clade consisting two African families. The divergence time which was calculated between the African and Asian notopterids belonged to the early Cretaceous at the time of India–Madagascar segregated from the African part of Gondwanaland. Therefore, on the basis of molecular record, approximated time of divergence was in disagreement with the new dissolution model. The Asian Notopterids were migrated into Eurasia on the Indian sub-continent at the time of the Cretaceous to the Tertiary.

Satoh et al. (2016) conducted a comparative assessment the mitogenomes of 250 species of fish. The diversified structural aspects of the mitogenomes of fish and the genes were encoded and were unravelled. The attained outcomes would be vital for the analysis of nucleotide sequences. Finally, this research showed the importance of comparisons for analysing the mitogenome structure.

CHAPTER III

MATERIALS AND METHODS

The complete mitochondrial genome of *Chitala chitala* was retrieved from NCBI GenBank database (GenBank Accession no. ON764424). The complete mitogenome was previously sequenced and deposited to NCBI by the respective authority of the Department of Biotechnology, Sher-e-Bangla Agricultural University, Dhaka, Bangladesh. The complete mitogenome sequences of other seven (07) species were also retrieved (Table 1). Then the sequences including 13 protein coding genes (PCGs) were analysed by different bioinformatic tools. The information of the eight fish species, their mitogenome sequences including the PCGs is presented in tables 1 and 2.

Table 1. Fish species and mitochondrial genome used in this study.

Sl. no.	Name of fish species	GenBank accession no.	Reporting country
1	*Chitala chitala*	ON764424	Bangladesh
2	*Chitala ornate*	AP008923.1	Japan
3	*Chitala lopis*	AP008922.1	Japan
4	*Chitala blanci*	AP008921.1	Japan
5	*Notopterus notopterus*	AP008924.1	India
6	*Papyrocranus congoensis*	AP008926.1	Japan
7	*Xanomystus nigri*	AP008927.1	Japan
8	*Osteoglossum biccirhosum*	AB043025.1	Japan

Table 2. The list of 13 protein coding genes with acronyms.

Sl. no.	Full name	Acronym
1	NADH dehydrogenase subunit 1	ND1
2	NADH dehydrogenase subunit 2	ND2
3	Cytochrome c oxidase subunit I	COX1
4	Cytochrome c oxidase subunit II	COX2
5	ATP synthase F0 subunit 8	ATP8
6	ATP synthase F0 subunit 6	ATP6
7	Cytochrome c oxidase subunit III	COX3
8	NADH dehydrogenase subunit 3	ND3
9	NADH dehydrogenase subunit 4L	ND4L
10	NADH dehydrogenase subunit 4	ND4
11	NADH dehydrogenase subunit 5	ND5
12	NADH dehydrogenase subunit 6	ND6
13	Cytochrome b	CYTB

3.1.Skewness analysis

Base composition for 13 PCGs(protein coding genes) of the mitogenome of *C. chitala* was analysed and the skewness (Perna et al., 1995) of their DNA sequences was calculated by using the formula where the number correlative of As is to Ts (AT skew = [A - T]/[A + T]) firstly (Sima et al., 2013) and the number correlative of Gs is to Cs (GC skew = [G - C]/[G + C]) lastly (Sima et al., 2013). Skewness because of any particular nucleotides is ascribed to differential mutational pressures originated from the mitochondrial DNA unbalanced replication (Yang et al., 2018) CAIcal is a web server which is available for free at http://genomes.urv.es/CAIcal, was used to estimate important computational gene parameters. Some of the gene parameters are codon adaptation index (CAI) values (Sharp et al., 1987), the effective number of codon (ENc) values (Wright et al.,1990) or G plus C percentage. Assessment of codon adaption index (CAI) is generally applied to enumerate codon usage bias in the expression of gene (Sharp et al., 1987). In general, the values of CAI limit from 0 to 1, where upper values indicate a upper percentage of codons which were the most available (Mazumder et al., 2015).

3.2.The values of effective number of codon (ENc)

Effective number of codon (Fuglsang et al., 2004) was counted to identify the independent codon bias which were synonymous found in a genomes' protein coding genes. It indicates to what extent all 61 codons are inclined to express a gene. A genes' expressivity depends on the efficient execution of codon usage (Gouy et al., 1982) for example, most available codons and codons interacting to numerous tRNA species in the cells are preferred in case of highly expressed genes (Ikemura et al.,1985) When the effective number of codons directs toward 20 it indicate that the respective gene is biased with its codons, while an unbiased gene has the proximity toward 61.

3.3.Codon usage and RSCU assessment

The frequency of codon usage and the values of relative synonymous codon usage (RSCU) were calculated by applying MEGA11 (Kumar et al., 2016). The estimation of RSCU value of a codon refers to the proportion of a codons' observed incidents to that of the expected usage frequency under the assumption that all codons which are synonymous for a definite amino acid partake uniformly (Sharp et al., 1986; Yengkhom et al., 2019).

The standard value of RSCU for each codon equals to one whereas altogether an amino acids' synonymous codons participate without any preferences. The codons which have RSCU values >1.6, are considered being "over-presented" whereas the codons which have RSCU values <0.6, are considered being "under-presented" (Gupta et al., 2001; Ma et al., 2018).

3.4.Selection pressure and mutation rate analysis

For analysing the evolutionary processes of DNA sequence of protein-coding genes, a powerful computational tool is applied to evaluate non-synonymous (substitutions that replace an amino acid by another) and the rates of synonymous substitution in the sequences that were coding, referred as Ka and Ks, or dN and dS, respectively (Hurst et al, 2002; Zhang et al., 2006). Ka or dN refers to non-synonymous substitutions for each non-synonymous site, and Ks or dS refers to synonymous substitutions for each synonymous site (Zhang et al.,2006). However, the Ka/Ks proportion (or ω, dN/dS) is frequently applied as an indicator of selective strength for the DNA sequences' evolution, with $\omega > 1$ implying positive or adaptive selection, $\omega < 1$ implying negative selection, and ω near 1 implying mutation which was neutral. In this study, Ka/Ks ratios were calculated by using an online based Synonymous non-synonymous Analysis Program (SNAP v2.1.1) (www.hiv.lanl.gov) website (Korber et al., 2000) where the protein coding genes' sequences were uploaded after aligning by Clustal W.

3.5.Sequence Analysis for Secondary Structure

The predicted OL and control region were identified by the analogy with the homologous sequences of other notopterids and proposed secondary structure. The secondary structure of the predicted OL was analyzed with the Mfold v.3.2 program on the (http://mfold.bioinfo.rpi.edu/) website with settings which were default (Zuker 2003).

3.6.Phylogenetic Analyses

To establish evolutionary relationships among the related species, the complete mitogenomes of *C. chitala* and related seven (07) species were retrieved from NCBI GenBank. The phylogenetic tree was created including the sequences which were collected from the control region (D-loop).

The Clustal W software was utilized to perform the alignment of individual genes between multiple species and excluded the start and stop codon. The *Osteoglossum bicirrhosum* (GenBank accession number: AB043025.1) was used as the out-group to determine the root of phylogenetic tree. The maximum Likelihood (ML) method was applied and the ML tree was constructed with 1000 replicates of bootstrap.

3.7.Estimation of Times of Divergence

The times of divergence of the Series Eupercaria was calculated by applying MEGA 11 (Tamura K, et al., 2021) with Molecular Clock Hypothesis in the method of RelTime Maximum Likelihood.

CHAPTER IV
RESULTS AND DISCUSSION

The mitogenome sequences of *Chitala chitala* and other seven (07) species were analyzed by different software and tools, the results of which were discussed below:

4.1. Composition of mitochondrial genome

The mitogenome sequences of *Chitala chitala, C. ornata, C. lopis , C. blanci, N. notopterus, P. congoensis, X. nigri and O. bicirrhosum* were 16,248, 16,164, 16,177, 16272, 16030, 16823, 16635, 16006 bp in length, respectively(Table 3). The mitogenomes contained 13 PCGs (ATP6, ATP8, CYTB, COX1, COX2, COX3, ND1, ND2, ND3, ND4, ND4L, ND5 and ND6), 2 rRNA genes (12S rRNA and 16S rRNA), 22 tRNA genes and one D-loop region (Figure 1). Among these genes, 8 tRNA genes (Gln, Ala, Asn, Cys, Tyr, Ser, Glu, and Pro) and ND6 were encoded on the light strand and the other 14 genes those were situated on the heavy strand.

The sizes of protein-coding genes (PCGs), ribosomal RNAs (rRNAs), transfer RNAs (tRNAs) and D-loop or control region for the eight species were compared (Figure 1). The maximum length diversification was detected in D-loop, and its' variation was regarded as the responsible for the differences of whole mitogenomes length. Besides, the D-loop length of *P. congoensis* was the highest resulted in its longest mitogenome.

Table 3. Total length of retrieved mitogenome sequences of *Chitala chitala* and other seven species.

Sl. no.	Name of fish species	Mitogenome length (bp)
1	*Chitala chitala*	16248
2	*Chitala ornata*	16164
3	*Chitala lopis*	16177
4	*Chitala blanci*	16272
5	*Notopterus notopterus*	16030
6	*Papyrocranus congoensis*	16823
7	*Xanomystus nigri*	16635
8	*Osteoglossum biccirhosum*	16606

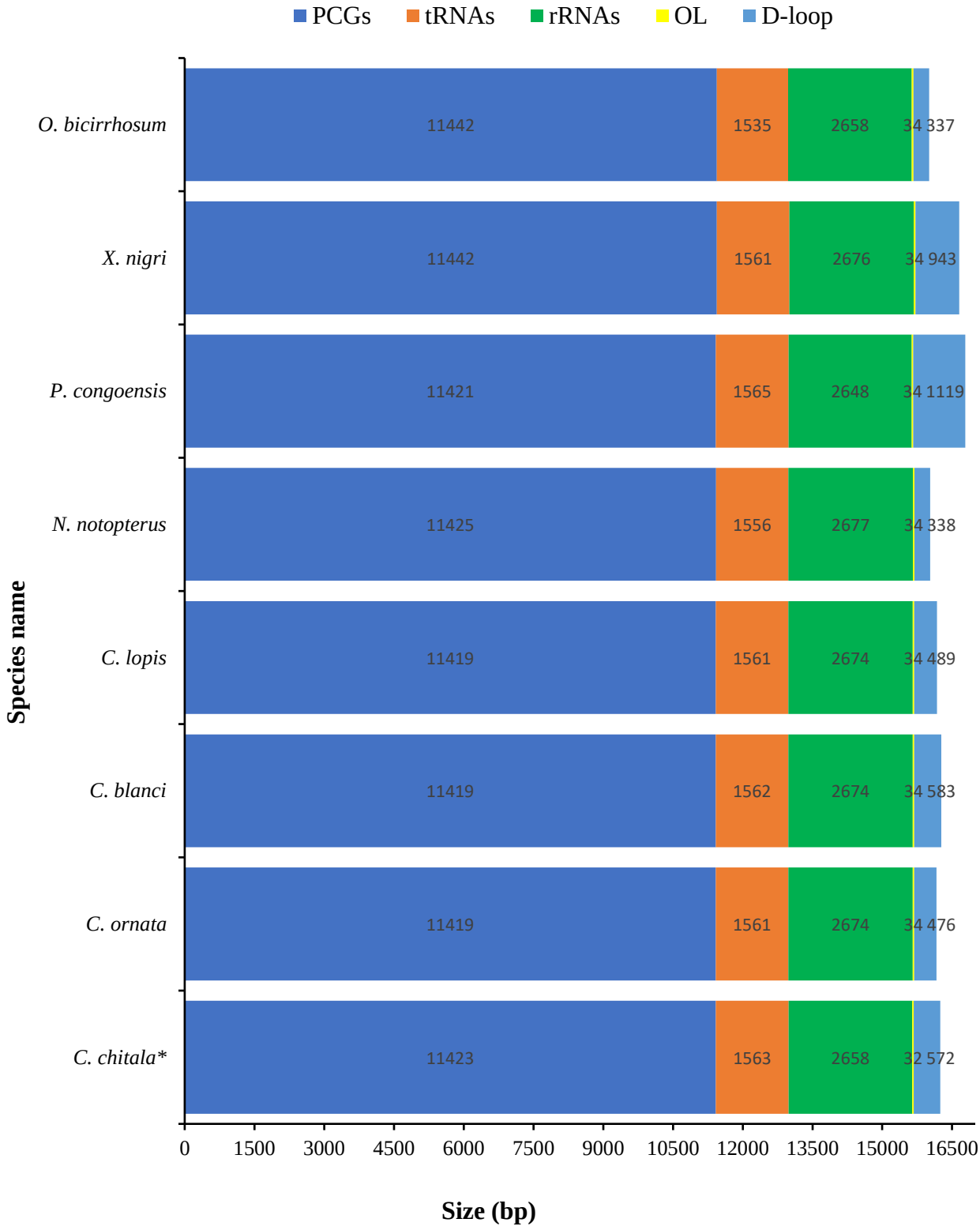

Figure 1. Relative size of PCGs, tRNAs, rRNAs, OL, control regions among 8 species (* present study)

The 13 PCGs of *Chitala chitala, C. ornata, C. lopis, C. blanci, N. notopterus, P. congoensis, X. nigri and O. bicirrhosum* were 11,423, 11,419, 11,419, 11,419, 11,425, 11,421, 11,442 and 11,442 bp in length respectively (Figure 1).

There were 22 tRNA genes observed in mitogenomes of *Chitala chitala, C. ornata, C. lopis, C. blanci , N. notopterus, P. congoensis, X. nigri and O. bicirrhosum* with the total of 1535-1564 bp in lengths (Figure 1).

The small and large rRNA genes were recognized in the H-strand in mitogenomes of *Chitala chitala, C. ornata, C. lopis, C. blanci , N. notopterus, P. congoensis, X. nigri and O. bicirrhosum*, ranging from 2648 to 2677bp in length, respectively (Figure 1). For *C. chitala* small 12s rRNA was 956 bp and large 16s rRNA was 1702 bp. Defining the boundaries of rRNA genes seems to be more difficult than the PCGs which had functional annotation features. Therefore, the boundaries of the genes could be inferred by assuming that there was no overlap or gap between contiguous genes.

In animal mitochondrial genes spacers were short, and can be used for evolutionary studies due to quickly changing ratio than gene regions (Hayward et al., 1997). There were several small intergenic spacers (IGS) region in *Chitala chitala, C. ornata, C. lopis, C. blanci , N. notopterus, P. congoensis, X. nigri and O. bicirrhosum.* The number and the size of intergenic spaces was one of the reasons for mitogenome length variation. Besides, the IGSs situated in the middle of tRNA-Asn and tRNA-Cys representing start signals for replication of L-strand (Yu et al., 2015) where OL range length variation of 32 bp to 34 bp (Figure 1).

Even though, the primary sequences of D-loop seem to play minor roles in regulatory function, as the region reveals wide variability across species even their relationship were close (Boore et al., 2006). In the present study, there was variation noticed od the D-loop lengths (Figure 1). The rapid variation of D-loop seems to provide some information for species evolution, but its internal mechanism needs more data and further examinations.

4.2.Protein Coding Genes (PCGs)

From the relative length of individual protein coding genes of 8 species it was observed that the thirteen protein coding genes had differences in their respective lengths. Firstly, for ND1 the range were from 969 to 975 bp, for ND2 1045-1047 bp, for COX1 1548-1566 bp, for COX2 688- 691 bp, for ATP8 168 bp, for ATP6 683-684 bp, for COX3 785-786 bp, for ND3 346-349 bp, for ND4L 297 bp, for ND4 1381 bp, for ND5 1838-1845 bp, for ND6 519-523 bp and for CYTB 1141 bp respectively(Figure 2).The encoding of 12 PCGs were on the H-strand, only the ND6 were encoded on the L-strand in the mitogenomes. So, there was not much of variation noticed in terms of size but there would be variation in terms of nucleotide bases.

There were overlapping regions also contained in the mitogenomes. In bony fishes, the overlapping of PCGs often intended that transcripts were partially shared between abutting regions, and most reading-frames overlaps were discovered in ATP8-ATP6, ATP6-COX3, ND5-ND6 and ND4L-ND4 (Sui et al., 2016). The discovery of the gene overlapping in the same region might suggest recent common ancestry and a putative genera-specific pattern (Montoya et al., 1983). The overlapping of the genes was one reason for mitochondrial genome compact, and the smaller mitochondrial genomes pass to offspring more frequently than the larger ones (Hogan et al., 2019). Nevertheless, selection was responsible for genomes size variation, relating to adapt new environment (Aguilar et al., 2019). The identical overlapping regions were detected in the featherback species, indicating they might have their own mechanism to cope with the environment.

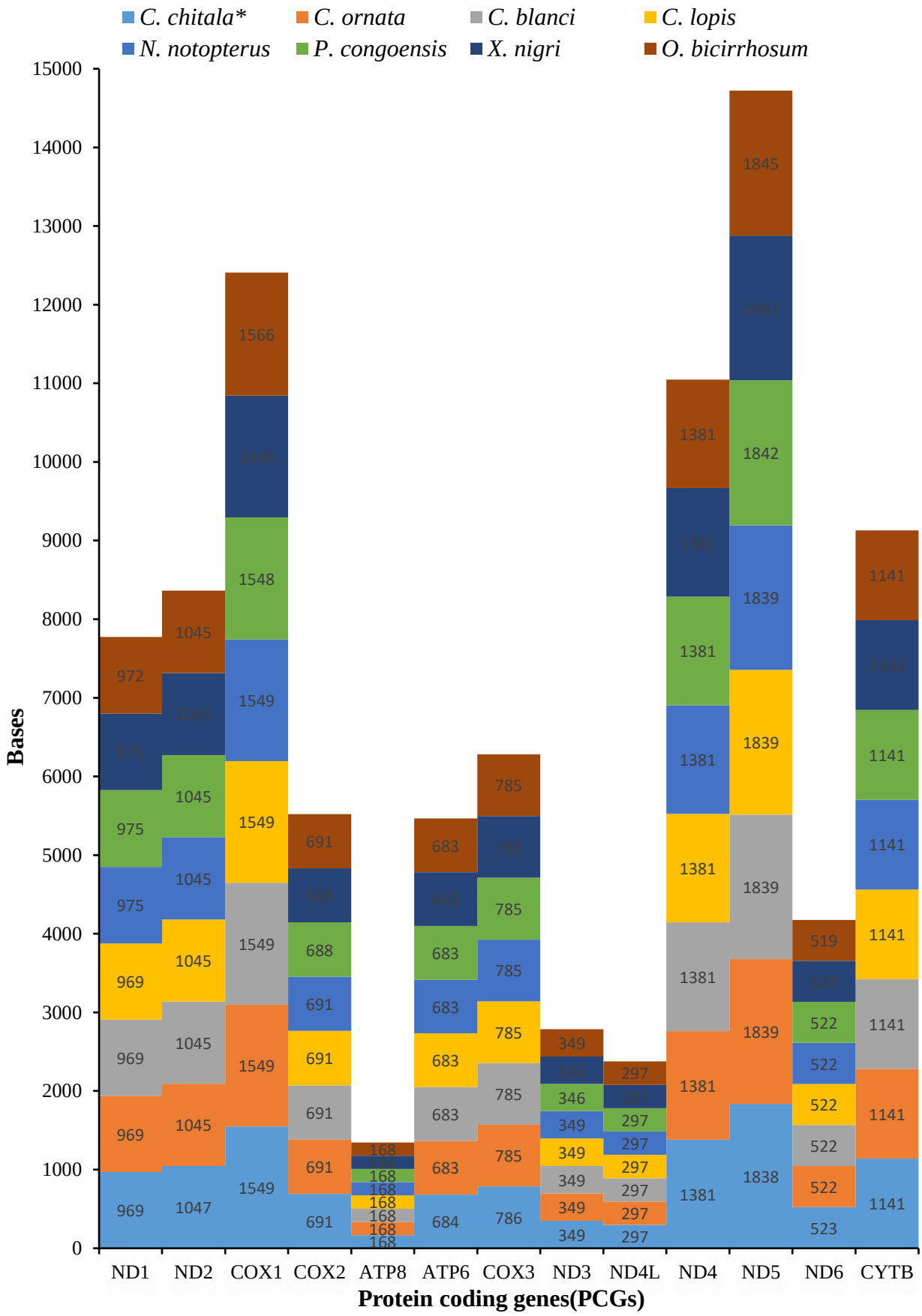

Figure 2. Relative length of individual protein coding genes of 8 species (* present study)

4.3. AT/GC-skewness

On the basis of the base composition of 13 PCGs of the mitogenome of *C. chitala*, the AT-GC-skews analysis showed an over-representation of A and C resulting a lower number of T and G bases, respectively, where there was slightly positive AT-skew (0.115) and slightly negative GC-skew (-0.225) (Figure 3). Except for ND6, all protein coding genes (PCGs) showed negative GC-skew, that indicated the higher occurrence of Cs. On the other hand, the COX1 and ND4L had slightly negative AT-skew (-0.03). Nucleotide base composition of each PCG was presented in the Appendix table 1.

The skewness of nucleotides might be ascribed to the equilibrium between mutation and selection pressures those are known to be affected by asymmetric mechanisms. By largely of mitochondrial DNA replication, repair and to a lesser extent of transcription (Fonseca et al., 2014) which ascertaining the functionality of related genes. AT-skewness and GC-skewness are considered as an important exponent to analyze strand asymmetry and the patterns of nucleotide composition of DNA sequences (Lü et al., 2019). The highest AT-skewness and highest GC-skewness value were observed in case of ND6, which was the only PCG exhibiting a positive value in the GC-skew curve and was common to other reports (Ruan et al., 2020; Mu et al., 2015). The genomes of mitochondria displayed compositional bias which were strand-specific and powerful as well (Asakawa et al., 1991). Therein heavy strand (H-strand) was rich in guanine and light-strand (L- strand) was poor in guanine (Anderson et al., 1981). In this analysis, ND6 which was located on the L- strand, showed larger fluctuation in AT/GC-skew value, suggesting that the selection and differential mutational pressure on this gene might be significantly different from other genes as reported in other fish species.

In the sequences of DNA, AT-skew and GC-skew were considered as potential indicator to measure strand asymmetry and the patterns of nucleotide composition (Lü et al., 2019). There were negative values found in majority of the AT and GC-skew of 13 protein-coding genes in

C. chitala, demonstrating base T and C were more abundant than A and G (Figure 3). In many cases, the prevalence of the GC-skew was larger than the AT-skew, and it was not statistically significant (Lü et al., 2019).

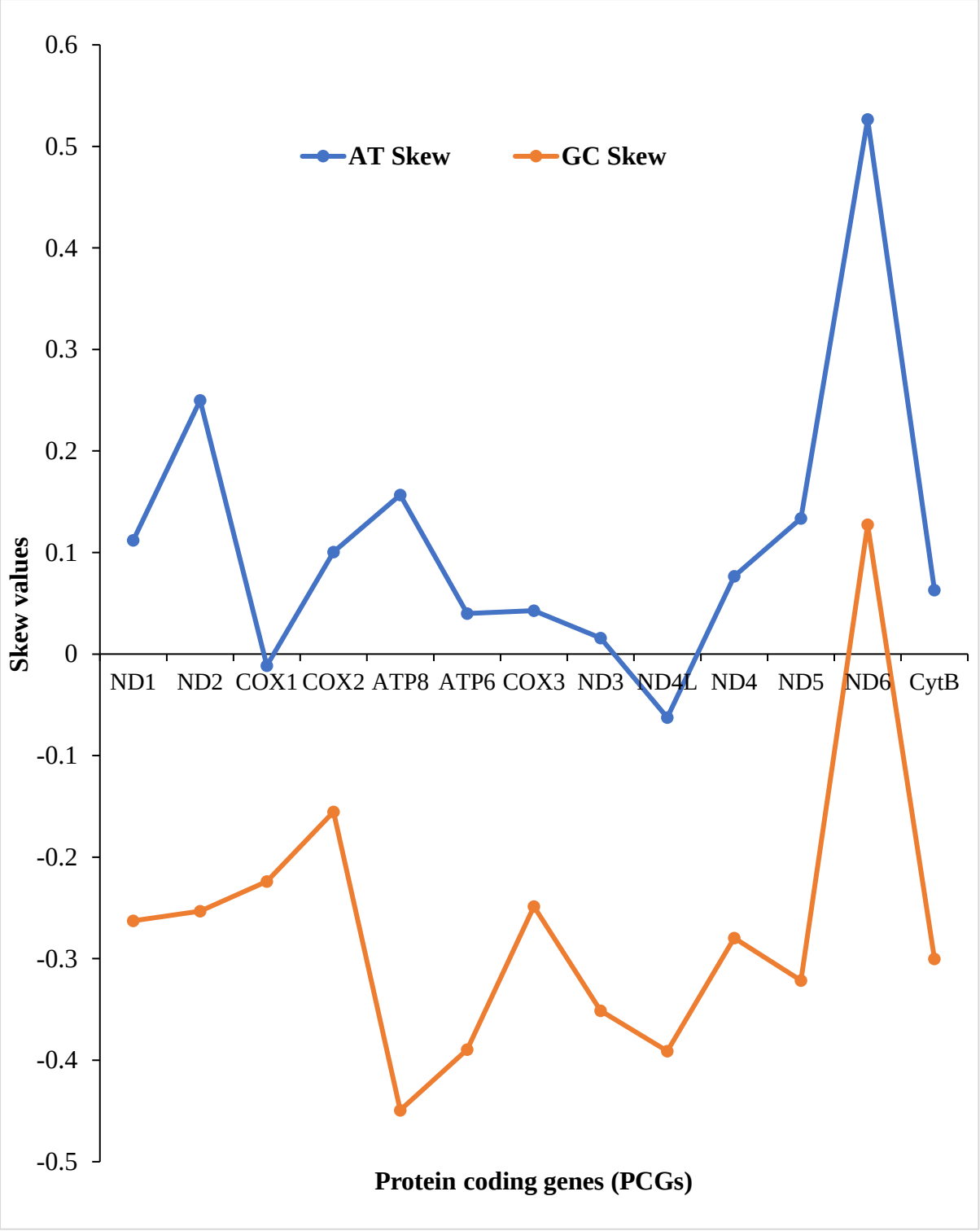

Figure 3. Graphical presentation of AT and GC Skew value of 13 PCGs for *C. chitala*

4.4.Gene parameters

There were three types of gene parameters measured for the eight species. The gene parameters were Codon adaption index (CAI), Percentages of GC (%GC) and Effective number of codon (ENc).The three gene parameters were measured for the thirteen protein coding genes.

Codon adaption index (CAI) of the protein coding genes of the eight studied fish ranged from 0.707 to 0.816 for *Chitala chitala,* from 0.578 to 0.658 for *C. ornata,* from 0.567 to 0.685*C. blanci ,* 0.593 to 0.643 for *C. lopis,* from 0.585 to 0. 685 for *N. notopterus,* from 0.575 to 0.658 for *P. congoensis,* from 0.576 to 0.660 for *X. nigri* and from 0.592 to 0.689 for *O. bicirrhosum.* Codon adaption index (CAI) of the protein coding genes of *Chitala chitala* ranged from 0.707 to 0.816 implied that they might have experienced slightly higher codon usage bias to be used in gene expression (Table 4).

Mazumder et al. (2015) reported that the level of gene expression was measured by CAI. In the study, it was evident that the gene expression level was more in *S. haematobium* than *W. bancrofti* due to high CAI values.

In this study, the values of codon adaption index (CAI) of the thirteen protein coding genes of *Chitala chitala* were ranged from 0.707 to 0.816 and these values of CAI were slightly higher than the other seven species implied that they might have experienced a higher codon usage bias to be used in gene expression (Table 4). Therefore, the CAI values were slightly higher than the other seven species as the fact was that the protein coding genes of *Chitala chitala* might be experienced higher level of gene expression then the rest seven species.

Table 4. CAI for all the protein-coding genes (PCGs) of 8 species' mitogenomes

PCGs	***C. ornata***	***C. blanci***	***C. lopis***	***N. notopterus***	***P. congoensis***	***X. nigri***	***O. bicirrhosum***	***C. chitala****
ND1	0.607	0.594	0.593	0.585	0.585	0.620	0.606	0.741
ND2	0.602	0.602	0.603	0.598	0.612	0.634	0.612	0.780
COX1	0.594	0.613	0.614	0.606	0.640	0.664	0.628	0.797
COX2	0.587	0.594	0.611	0.601	0.593	0.627	0.624	0.811
ATP8	0.578	0.567	0.609	0.593	0.575	0.576	0.614	0.816
ATP6	0.585	0.585	0.615	0.605	0.583	0.593	0.611	0.728
COX3	0.596	0.608	0.624	0.612	0.581	0.612	0.608	0.707
ND3	0.613	0.621	0.612	0.607	0.579	0.625	0.601	0.736
ND4L	0.637	0.637	0.609	0.599	0.576	0.612	0.592	0.757
ND4	0.642	0.623	0.622	0.619	0.587	0.618	0.607	0.745
ND5	0.632	0.632	0.634	0.626	0.639	0.626	0.638	0.731
ND6	0.658	0.685	0.669	0.685	0.658	0.660	0.687	0.728
CYTB	0.638	0.635	0.643	0.648	0.632	0.647	0.689	0.716

Table 5. ENc for all the protein-coding genes (PCGs) of 8 species' mitogenomes

PCGs	***C. ornata***	***C. blanci***	***C. lopis***	***N. notopterus***	***P. congoensis***	***X. nigri***	***O. bicirrhosum***	***C. chitala****
ND1	34.7	37.9	37	37.4	34.7	40.7	39.5	48.2
ND2	38.2	39.5	42.5	42.5	38.2	41.4	37.5	43
COX1	41.3	41.7	45.6	46.8	43.3	43.2	45.5	45.6
COX2	43.6	43.8	52.3	51.7	45.6	36.4	42.4	44.1
ATP8	45.4	45.7	57.4	54.3	43.2	34.6	34.6	43.6
ATP6	43.8	44.3	55.1	52.8	41.7	37.2	36.3	49.3
COX3	41.7	43.2	53.6	49.6	42.8	38	38.5	49.9
ND3	42.3	42.7	51.7	44.8	43.9	40.4	35.7	45.8
ND4L	41.5	41.5	50.4	41.2	44.7	44.2	33.8	40.8
ND4	40.9	41.1	46.8	40.7	43.9	42.6	37.3	42.7
ND5	40.5	40.5	37.1	39.8	43.3	41.1	39.9	41.9
ND6	41.3	41.3	39.2	39.3	38.2	37.9	39.1	41.2
CYTB	39.7	40.7	38.2	38.4	36.5	35.3	36.2	40.3

Effective number of codon (ENc) of the protein coding genes of the eight studied fish ranged from 40.3 to 49.9 for *Chitala chitala*, from 34.7 to 45.4 for *C. ornata*, from 37.9 to 45.7 *C. blanci* , 37 to 57.4 for *C. lopis*, from 37.4 to 54.3 for *N. notopterus*, from 34 .7 to 45.6 for *P. congoensis*, from 35.3 to 44.2 for *X. nigri* and from 33.8 to 45.5 for *O. bicirrhosum.* Effective number of codon (ENc) for *Chitala chitala* was varied from 40.3 to 49.9 (Table 5).

Effective number of codon (ENc) for *Chitala chitala* was varied from 40.3 to 49.9 for the thirteen protein coding genes (Table 4). When the value of the effective number of codons was close to 20, it referred that the gene was biased with its codons. On the contrary, the value of the effective number of codons directed toward 61 for the unbiased gene. As the value of the effective number of codons for *Chitala chitala* was varied from 40.3 to 49.9, so the protein coding genes showed average biasness with the codons (Table 5).

Table 6. Percentages of GC for all the protein-coding genes (PCGs) of 8 species' mitogenomes

PCGs	***C. ornata***	***C. blanci***	***C. lopis***	***N. notopterus***	***P. congoensis***	***X. nigri***	***O. bicirrhosum***	***C. chitala****
ND1	46.1	46.1	46	43	42.4	45.1	44.5	46.5
ND2	43.8	44.8	45.4	44.5	42.9	46.3	43.1	44.2
COX1	41.5	43.6	44.7	45.3	43.5	47.9	41.3	42.5
COX2	39.7	41.4	42.3	43.6	41.4	42.8	40.4	40.8
ATP8	37.5	37.5	41.7	43.5	38.7	41.1	42.3	39.3
ATP6	40.2	39.2	43.1	44.1	40.5	43.6	43.2	41.5
COX3	42.8	41.6	43.8	44.7	42.3	47.8	41.7	46.3
ND3	44.9	45.5	45.4	45.5	42.9	44.7	44.4	44.7
ND4L	46.8	45.7	46.8	46.8	43.8	47.8	45.5	46.1
ND4	43.4	44.3	44.5	40.8	41.6	46.1	43.8	45.3
ND5	41.7	41.7	41.9	39.1	40.1	43.9	42.9	43.9
ND6	42	42	40.8	42.3	41.4	48.9	42	43.4
CYTB	40.7	41.3	39.3	41.6	39.8	45.2	41.3	41.7

Percentages of GC of the protein coding genes of the eight studied fish ranged from 39.3 to 46.5 for *Chitala chitala*, from 37.5 to 46.8 for *C. ornata*, from 37.5 to 46.1 *C. blanci* , 40.8 to 46.8 for *C. lopis*, from 39.1 to 45.5 for *N. notopterus*, from 38.7 to 43.8 for *P. congoensis*, from 41.1 to 48.9 for *X. nigri* and from 41.3 to 45.5 for *O. bicirrhosum*. Therefore not much variation was observed here (Table 6).

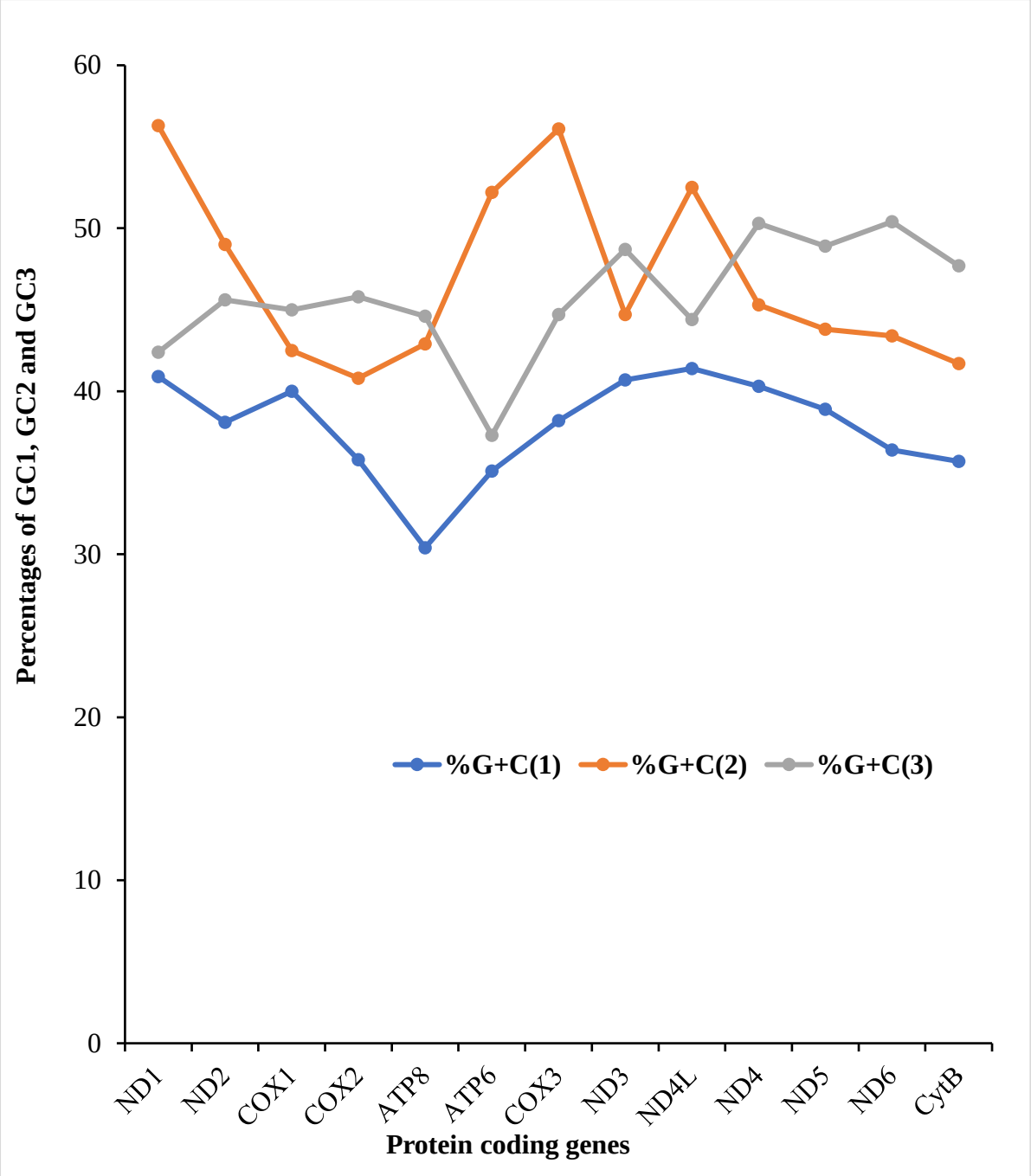

Figure 4. Percentages of GC content at the codon positions 1st, 2nd and 3rd of 13 PCGs of *Chitala chitala*

Percentage of GC bases located on the codon positions 1st, 2nd, 3rd were shown in (Figure 4), where % GC1 was lower than % GC2 and % GC3 across all the PCGs and % GC3 was highest. Jenkins et al. (2003) suggested that mutation pressure, genetic drift and natural selection were the major elements from which the codon usage biases were influenced (Wu et al., 2015). On the other hand, GC contents at the third codon position (Wu, et al., 2015), gene expression levels and gene length also were related to the codon bias (Gustafsson et al., 2004). The main evolutionary force led to high content of A + T or G + C was the mutation pressure in animals. Comparing with the low GC content gene, higher GC content of third codon position seemed easier methylated and caused mutations (Tatarinova et al., 2010). In the current study, % GC3 was highest implied that the third codon position had the higher GC content so it might be easier methylated. As a result, mutation might be occurred (Figure 4).

4.5.Codon usage patterns

The amino acids Serine and Leucine were encoded by six different codons while rest of the amino acids were encoding by two or four codons likewise. Hydrophobic amino acid Pro (30.9 %) was utilized at the highest frequency followed by hydrophobic amino acids, Ser (29.2 %) and Leu (27.3 %), whereas the frequency of the hydrophilic amino acid, Glu (1.2 %) was the lowest among the 13 PCGs (Figure 5A). The most frequently used codon was AAT (15.7 %, Asn) followed by CCT (15.6 %, Pro) whereas, the rarely used codons were GCG (Ala), GAA (Glu) and GTG (Val).

The RSCU value of all amino acids in the three species were not equal to 1, implying that the usage of each amino acid had varying degrees of bias (Figure 5B). The analysis of relative synonymous codon usage (RSCU) exposed a total of 22 values of RSCU with higher than 1.0 (Figure 5B). Several codons, AAA(Lys), CCT(Pro), ACT(Thr), CCT(Leu), AGC(Ser) and GTT(Val) were over-presented whereas the codons, GCG(Ala), CAG(Gln), AAG(Lys), CCG(Pro), TCG(Ser), ACG(Thr) and GTG(Val) were under-presented for coding the PCGs of the *C. chitala* mitogenome. The pattern of codon usage is exploited for those codons which

are more quickly and efficiently translated into protein by the cellular machineries of an organism. Nevertheless, translational selection is demonstrated in case of genes those are expressed highly by engaging the models of codon usage. The codons that are preferred usually occur at higher frequencies in the sequence of highly expressed genes than that in weakly expressed genes (Jia et al., 2008).

The value of RSCU was also used to measure mitochondrial gene codon usage. When the RSCU value was 1, it indicated that the frequency of use of codons had no different with other degenerate codons; when the value of RSCU greater than 1, it represented the codon was employed more repeatedly (Sharp et al., 1986; Mazumder et al., 2015). The RSCU value of all amino acids in the three species were not equal to 1, implying that the usage of each amino acid had varying degrees of bias (Figure 5B). The biases of codon usage were significant in the mitochondrial genomes of different species, and it made the gene under different selection pressure and could predict the gene function (Sharp et al., 1994).

The value of RSCU refers to how many times a specific codon is noticed in compare to how many times that codon would be noticed for an identical synonymous codon usage. When there is in the absence of any codon usage bias the value of RSCU is 1.00. In case of a codon uses less repeatedly than anticipated then the values of RSCU is less than 1.00, on the contrary codons use more repeatedly than anticipated will have the values of RSCU more than 1.00.

Islam et al. (2022) reported that a sum of 25 RSCU values more than 1 were identified. Moreover, 12 codons were revealed as “over-presented” that implied to assign in highly expressed genes for efficient protein synthesis through translational selection process. Yang et al. (2018) reported 30 RSCU values in a cottoid fish. In the current study, a total of 22 RSCU values were found with the values of greater than 1. Furthermore, 6 codons were disclosed as “over-presented” that implied to engage in highly expressed genes for efficient protein synthesis via translational selection.

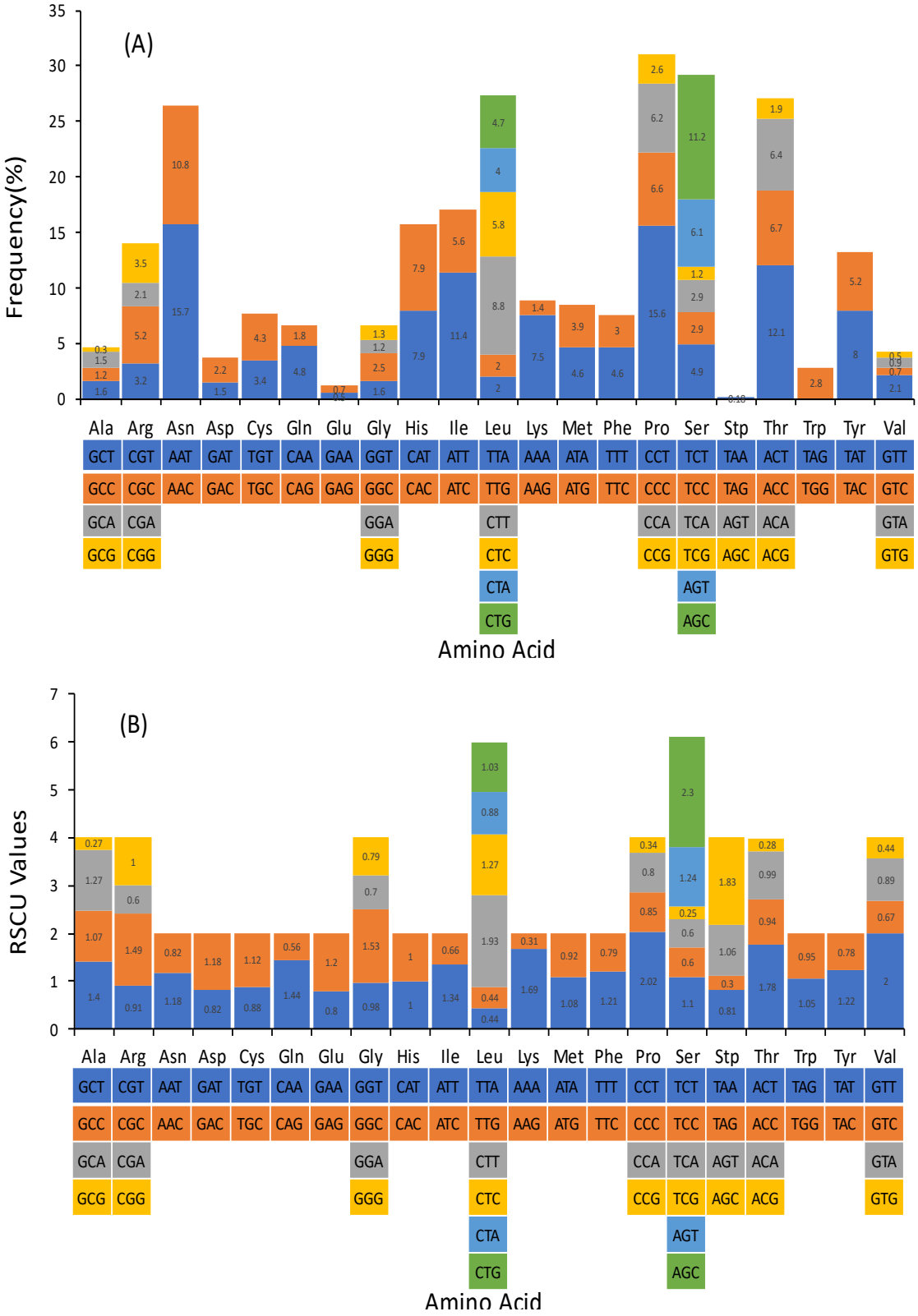

Figure 5. Codon usage frequency values (A) and relative synonymous codon usage (RSCU) values (B) of the mitochondrial genome of *Chitala chitala* (Reversed complementary sequence was applied for ND6 gene as it was encoded on the L-strand and incomplete stop codons were not included).

4.6.Evolution rates of PCGs

The sequences of the protein-coding genes of *C. chitala* mitogenome were retrieved then the rates of evolution were estimated referring to the mitogenomes of its phylogenetic relatives, *Chitala ornata* (GenBank accession no.: AP008923.1) and *Notopterus notopterus* (GenBank accession no.: AP008924.1). The estimated Ka/Ks ratios also known as dN/dS ratios between *C. chitala* and *C. ornata* ranged from the lowest value (0.01) for ATP6 gene to the highest value (0.996) for ND5 gene whereas, the values between *C. chitala* and *N. notopterus* ranged from the lowest value (0) for ND1 gene to the highest value (0.828) for CYTB gene (Figure 6). The ND5 gene showed the higher Ka/Ks ratios comparing to that of other protein coding genes among the three fish species.

In molecular studies, to calculate selective pressure and evolutionary relationships of species the proportion of non-synonymous substitution (Ka) and synonymous substitution (Ks) is a conventional exponent (Zhu et al., 2017; Jia et al., 2020). In case of Ka/Ks < 1 was referred as purifying or negative selection, Ka/Ks = 1 was referred as neutral mutation, and Ka/Ks > 1 was referred as positive selection (Zhu et al., 2017; Wang et al., 2022). Since Ka/Ks values less than 1, all the thirteen protein-coding genes were suffered from negative or purifying selection (Figure 6). On the contrary, the outcome was different from deep-sea fishes, where there were demonstration of positive selection from many fishes and similar signals with ND4L and ND5 genes excluded (Shen et al., 2019; Jia et al., 2020). It was among other facts that the difference in their habitats. There were random genetic drift and the pressure of mutation which linked to the environment, on which the evolution of genome being dependent on.(Schaack et al., 2019). In the state of deficiency of oxygen, lack of food, absence of sunlight and excessive cold, the fishes of deep-sea dwelled upon whereas the featherback species inhabited in freshwater(Shen et al., 2019; Jia et al., 2020). Nielsen et al. (2005) suggested, the selection which was positive generally connected with the acclimatization of new conditions of the niche and the improvement of neo operations.There was demonstration of undergoing negative or purifying

selection from the Ka/Ks values in notopteridae species. It was suggestive that their genetic function could not be altered by the variation of the environment.

ND5 and CYTB genes had high Ka/Ks values across *C. chitala* mitogenomes comparing to the rest of the genes, whereas ATP6 and ND3 genes were among the fewest (Figure 6). For the reason of repairing mechanisms of DNA, less rate of mutation inferred to happen on genes those were expressed immensely (Barrientos et al., 2002; Jia et al., 2020). On the comparison to other genes, the ATP6 and ND3 showed low Ka/Ks illustrated less rate of mutation. Which signified that they might have greater level of expression (Jia et al., 2020).

Yang et al. (2018) reported that among all of the 13 PCGs of three croakers were investigated for mutation rates in a study, where the Ka/Ks proportion expanded from 0.0100 to 0.2714. In that study, the Ka/Ks ratios indicated that the protein coding genes of those three fish species evolved under strong purifying selection. Lynch et al. (2006) suggested, the process of evolution of a DNA sequence is driven by forces that were not adaptive along with the pressure of mutation and genetic drift. In spite of the fact that, under purifying selection, it is difficult to establish any mutations driven demerits, the selection pressures posed differently among the genes and are supposed to evolve in different ways.

Nielsen et al. (2005) reported that there were two strengths that were not adaptive, which were the pressure of mutation and random genetic drift. The basic features of genome evolutional though working limitations imposed burdens on the events of mutation were ascertained by them (Shi et al., 2016). As a result, the disadvantages combined with mutation were hard to ordain undergoing the selection that was negative (Shi et al., 2016). Long-term stability was maintained by the selection processes of the biological structure. It was indicated by the Ka/Ks ratio that the evolution of the functional genes was undergoing extreme negative selection. That referred, natural selection was against deleterious mutations containing selective coefficients

those were negative. The pressures of selection varied among the genes. These were apparently various ways to evolve.

Islam et al. (2022) reported that the values were near one of the Ka/Ks ratio between *C. reba* and *L. bata* was which intended to mutation that was neutral in ND1 gene of *C. reba* implying no effect to its survival or reproduction. Outside of ND1, COX2 and ND5, the rest values of Ka/Ks ratios were less than 0.1. However, the mutation occurred in these PCGs of *C. reba* was controlled by purifying or negative selection could easily be thought.

In the current study, the ratio of Ka/Ks between *C. chitala* and *C. ornata* was close to one which meant neutral mutation in ND5 gene of *C. chitala* implying neither beneficial nor detrimental to its survival or reproduction. Except for COX2, ND5 and CYTB all the Ka/Ks proportional values were fewer than 0.1. Though, it can be considered that the mutation occurred in these PCGs of *C. chitala* was driven by purifying or negative selection. Negative selection implies the abundance of more synonymous mutations than non-synonymous mutations where ancestral state and more varied gene pool is essential to conserve for appropriate functions of encoded proteins. If negative selection is unable to remove the harmful mutations, an accumulation of deleterious mutations occurs (Kimura et al., 1968). Then selection occurs that reduce the number of surviving offspring. Thus, the possibility of extinction will increase. Climate changes and habitat alterations are responsible for threatening of those species under strong negative selection.

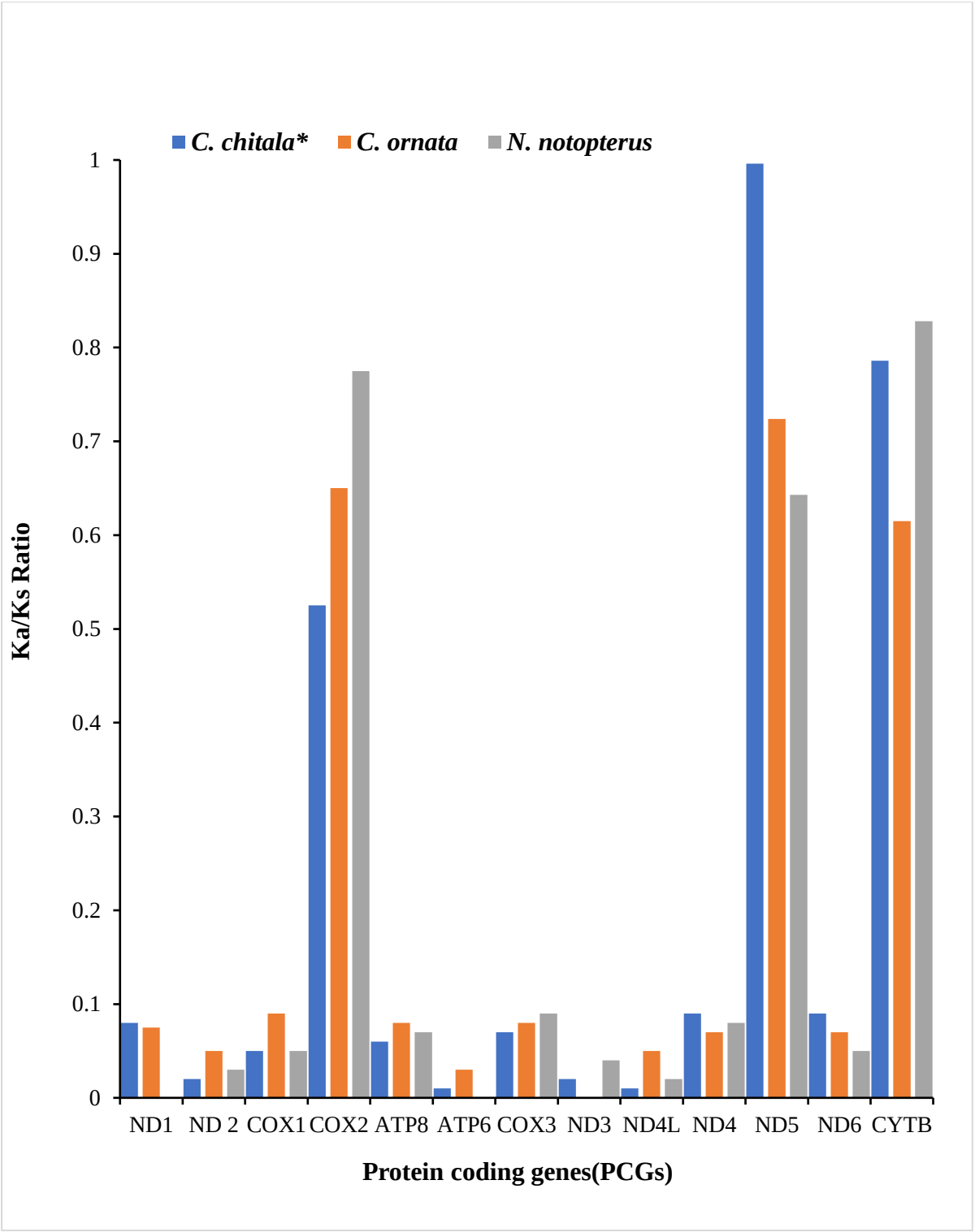

Figure 6. Selective strength of protein coding genes of *C. chitala* mitogenome comparing that of *C. ornata* and *N. notopterus*. For every protein-coding genes, the ratio indicates the rate of non-synonymous substitutions to the rate of synonymous substitutions (Ka/Ks). (* present study)

4.7.Phylogenetic relationship analysis

By using control regions, the construction of the phylogenetic relationship of *C. chitala* with other notopteridae fishes was done for the mitogenomes of the subfamily Notopteridae by MEGA 11. The mitogenome of the studied species, *C. chitala* (GenBank accession no. ON764424) showed 99% similarity of sequence with *C. blanci* (GenBank accession no. AP008921.1), followed by 88% identity with *C. lopis* (GenBank accession no. AP008922.1). However, the phylogenetic tree showed that there was formation of monophyly of four *Chitala* spcies and *Notopterus notopterus* with underlying bisection other three species. The root of the phylogenetic tree on the outgroup *Osteoglossum biccirhosum* showed that all of them were evolved from a common ancestor (Figure 7).

By using the Maximum Likelihood method and Tamura-Nei model, the history evolution was supposed (Tamura et al., 1993). For illustrating the evolutionary history of the taxa resolved the bootstrap consensus tree inferred from 1000 replicates (Tamura et al., 2021) was considered. In less than 50% bootstrap replicates, branches resembling to divisions reformed were deteriorated. In the replicate trees' percentile, the connected taxa clustered together in the bootstrap test 1000 replicates) were around the branches displayed (Tamura et al., 2021). By applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model inaugural tree(s) for the investigative search were attained automatically, and then choosing the topology with upper value of log likelihood (Imre et al., 2020). There were 8 nucleotide sequences included in this exploration. The positions of codon involved were 1st+2nd+3rd+Noncoding. In the last dataset, there were a total of 4135 positions. These analyses were performed in MEGA11.

Inoue et al. (2009) suggested that previous phylogenetic researches (Hilton et al., 2003; Lavoué, et al., 2004) originally permitted the monophyly of Notopteridae and the underlying bisection of African and Asian notopterids. The suppositional phylogenetic relationships were compatible with the present phylogeny that was studied (Figure 7).

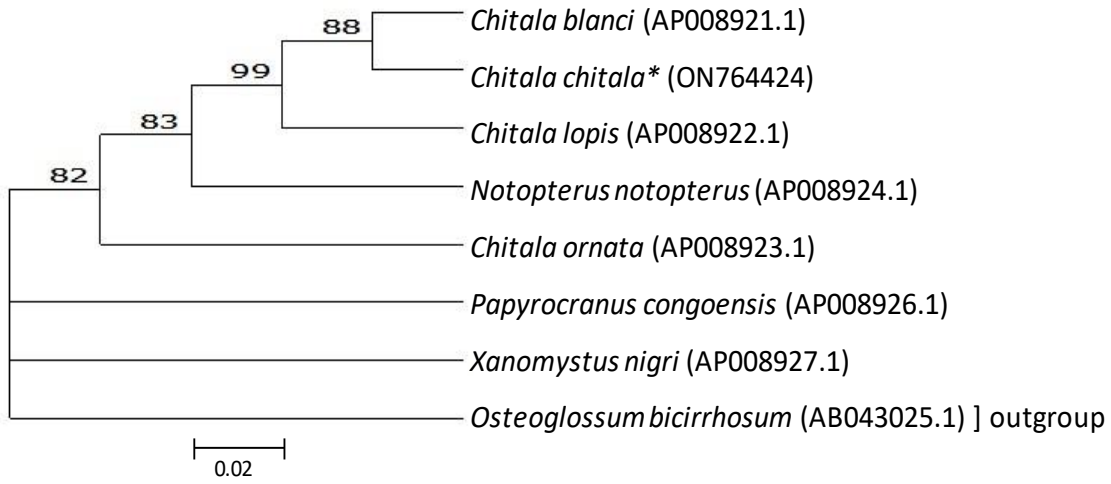

Figure 7. Phylogenetic tree of control regions (D-loop) of 8 species of the subfamily Notopteridae. The analysis of phylogenetic relationship was performed by MEGA 11 with Maximum Likelihood method and bootstrap replications of 1,000. For each species, GenBank accession number was given next to its scientific name. (* present study)

4.8.Estimation of Times of Divergence

The phylogenetic time tree of *C. chitala* with seven other notopteridae fishes was constructed using mitogenomes' control regions of 8 species from Notopteridae family by applying MEGA 11. The mitogenome of the studied species, *C. chitala* (GenBank accession no. ON764424) displayed divergence time 0.2 million years ago (MYA) whereas *Chitala blanci* (GenBank accession no. AP008921.1) was 0.1 MYA along with other species and *X. nigri* (GenBank accession no. AP008927.1) showed the highest divergence time 1.2 MYA. Estimation of times of divergence showed four *Chitala* species and *Notopterus notopterus* showed similar divergent times than other species. The root of the phylogenetic tree on the outgroup *Osteoglossum bicirrhosum* showed that all of them had been evolved from a common ancestor (Figure 8).

The time tree displayed by applying the RelTime method was created (Tamura et al., 2012; Hong et al., 2017). The Maximum Likelihood method and Tamura-Nei model were applied to enumerate divergence times for all ramifying points in the topology (Tamura et al., 1993). The approximate value of log likelihood of the topology was -10951.33. The portrayal of the tree to scale, with branch lengths measured in the relative number of per sites' substitutions (Hassan et al., 2020). 8 nucleotide sequences were used in this exploration. In the final dataset there were a sum of 4135 positions. In MEGA11, these analyses were managed (Tamura et al., 2021). Inoue et al. (2009) reported that among the Notopterids, there was a split between the African and Asian notopterids happened approximately 133 Mya and the longest split within the Asian notopterids was occured approximately 47 Mya (36–61 Mya).There was a negligible effect on the estimates of divergence time by alternating rttm from 419 to 472 Mya and high time from 900 Mya to 1800 Mya. And it was consistent with the present study (Figure 8).

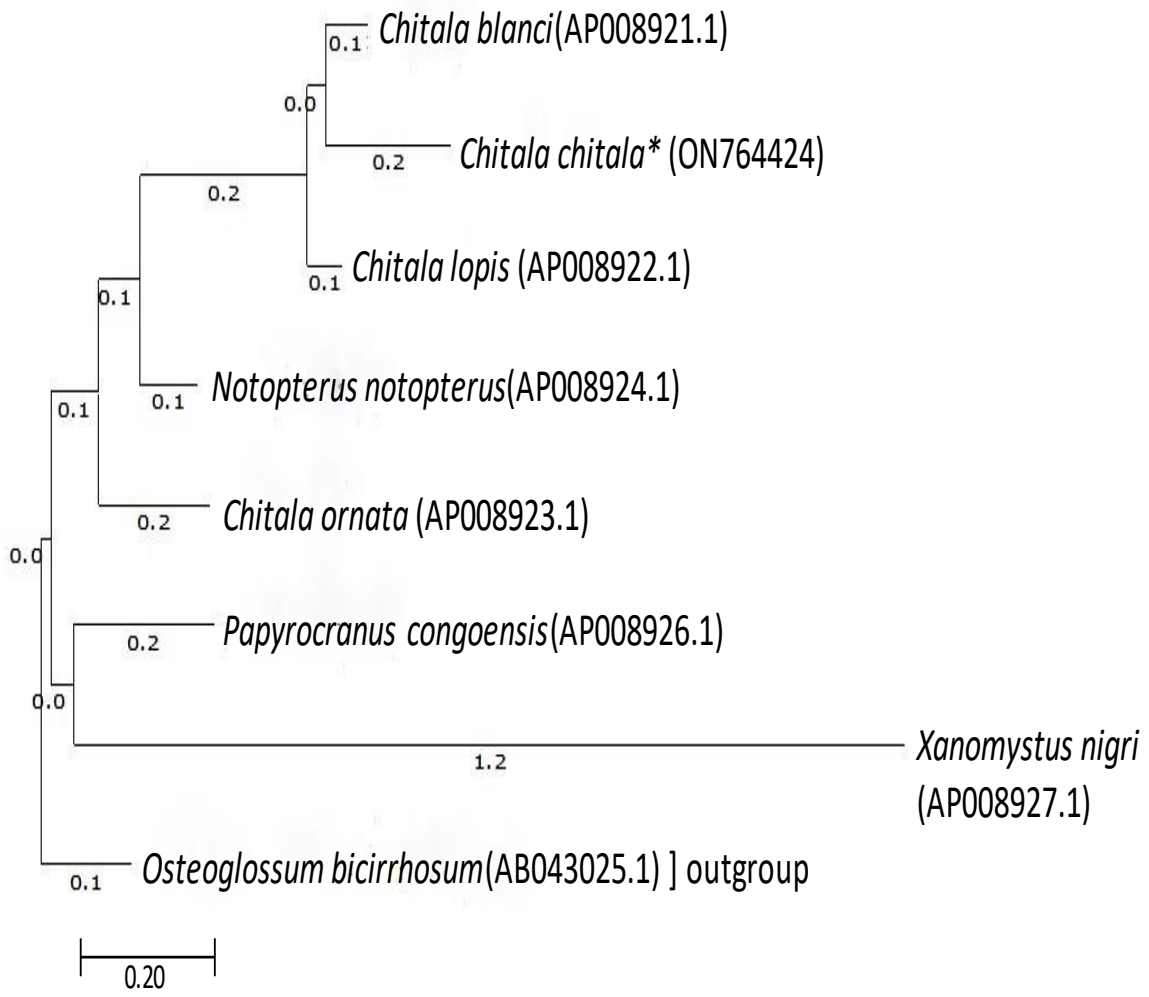

Figure 8. Time tree of *Chitala chitala* and its relatives under Notopteridae family using the control region(D-loop). The estimation of times of divergence was performed by MEGA 11 with Reltime Maximum Likelihood method. (* present study)

4.9.Secondary Structure of OL

There were similarity with various mitogenomes, *C. chitala* had no introns and occupied two non-coding regions, one origin of light strand (OL) and one D-loop control region. The OL region of *C. chitala* had a secuence of 32 nucleotides (5´-CCCGCCT TTTCCCGCCGCGGCGGGAGGAAGGT-3´), whereas *C. ornata* had 34 nucleotides (5'- CC TCCCGCCTTTTCCCGCCGCGGCGGGAGGAAGG-3') and *N. notopterus* had 34 nucleotides (5'-CCTCCCGCCTCCCCCCGCCTCGGCGGGAGGAAGG-3'), situated between $tRNA^{Asn}$ and $tRNA^{Cys}$. And it was positioned on the L-strand in a bunch of five tRNA genes (WANCY region). The sequence of OL region of *C. chitala* predictively had the ability of folding into a stem-loop secondary structure which was stable. It contained the loop of 14 bp and the stem of 11 bp (Figure.9).

The minimum folding free energy was -11.03 kcal/mole for *Chitala chitala*. The default 5% of this number was less than 1 kcal/mole, and so $\delta\delta G$, the free energy increment, was rounded up to 1 kcal/mole. Similarly, for *Chitala ornata* and *Notopterus notopterus* the minimum folding energy were -8.98 kcal/mole and -8.98 kcal/mole respectively.

The OL was always found in the intergenic region between two conserved genes (Brockman et al., 2012) and the three mitogenomes were discovered. The OL was usually observed in the middle of tRNA-Asn and tRNA-Cys in bony fishes. The main feature of OL was the structure which was secondary that was folded into a stable stem-loop (Liu et al., 2017).

Cheng et al. (2012) reported that the OL was located in a cluster of five tRNA genes (WANCY region) between $tRNA^{Asn}$ and $tRNA^{Cys}$ and was 36 bp in length. This region was predicted capable of folding into a stem-loop secondary structure with 13 bp in the stem and 10 bp in the loop which was stable. However there were similarities of this stem-loop structure to those of other Pseudosciaeniae species which had stems ranging in size from 12 bp to 13 bp and loops of about 10–14 nucleotides.

This characteristic stem-loop structure of the OLs with stem of 11 nucleotides and loops of 14-15 nucleotides (Figure 9) which was conserved and similar to those of other Pseudosciaeniae species which might have structural and functional importance in connection with the replication of L-strand (Hixson et al. 1986; Cheng et al. 2012). Their functional importance in combination with the origin of the replication of mitochondrial DNA was suggested by the conserved stem-loop structures in these mitochondrial genomes.

Cheng et al. (2012) also reported that a recognition sequence (5'-CGGCC-3') on the $tRNA^{Cys}$, at the base of the loop of OLs in the Pseudosciaeniae species, but the recognition sequence was not present in *Chitala chitala*, *C. ornata*, *N. notopterus* which might be due to different phylogenetic origin.

During the process of the formation of the secondary structure of OL of the mitochondrial DNA , for *Chitala chitala* the minimum folding free energy was -11.03 kcal/mole. The default 5% of this number was less than 1 kcal/mole, and so $\delta\delta G$, the free energy increment, was rounded up to 1 kcal/mole. It referred that during translation process, 11.03 kcal/mole free energy was consumed in the folding of DNA to create the secondary structure of OL for *Chitala chitala*. Likewise, for *Chitala ornata* 8.98 kcal/mole and for *Notopterus notopterus* 8.98 kcal/mole free energy were consumed respectively. So it was observed that *Chitala chitala* consumed more free energy than the others and that might be harmful for this species.

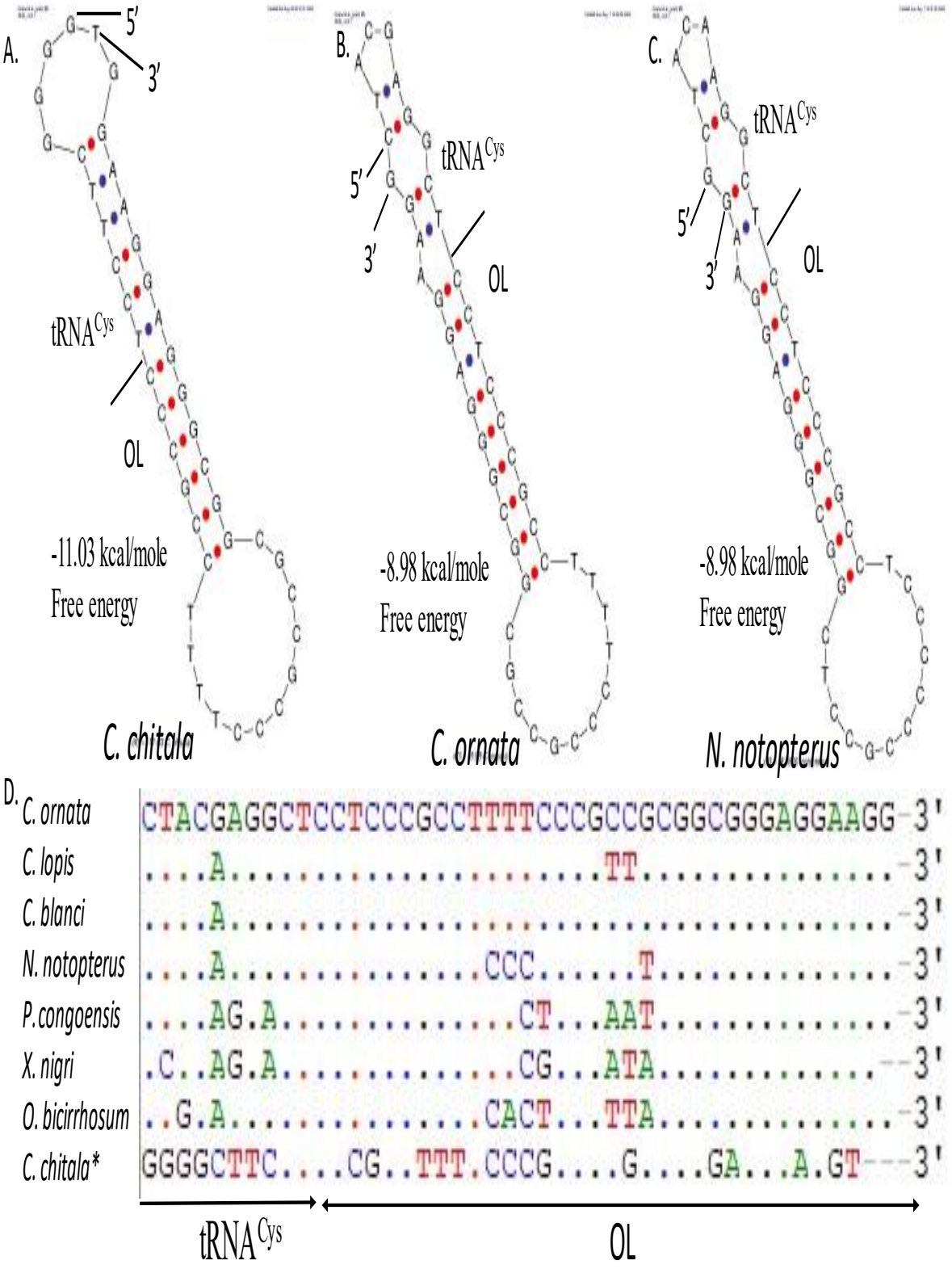

Figure 9. Predicted potential secondary loop-stem structure of OL of *C. chitala*, *C. ornata*, *N. notopterus* (A-C). Multiple sequence alignment of OL sequence of 8 species of notopteridae fishes (D). Reverse complementary sequences was used. Dotted positions are identical nucleotides (* present study)

CHAPTER V
SUMMARY AND CONCLUSION

The present research, was about the exploration for the sequence of mitogenome of *Chitala chitala*. The mitogenome was with a total size of 16,248 bp. The mitogenome composed of 13 PCGs, 22 tRNAs, 2 rRNAs and one control region. Most PCGs were started with ATG codon and terminated with TAA codon. The Codon adaption index (CAI) of the protein coding genes of *Chitala chitala* ranged from 0.707 to 0.816 referred that having an average codon usage bias in gene expression process. The amino acids Serine and Leucine were encoded by six different codons but the rest of the amino acids were encoding by two or four codons similarly. Among all the amino acids, the hydrophobic amino acid Pro (30.9 %) was utilized at the highest frequency then hydrophobic amino acids, Ser (29.2 %) and Leu (27.3 %). On the contrary, the frequency of the hydrophilic amino acid, Glu (1.2 %) was the lowest among the 13 PCGs. A total of 22 RSCU values (>1) were identified and 6 codons were revealed as "over-presented" that implied for codon usage bias to engage in highly expressed genes for efficient protein synthesis via translational selection. The proportion of Ka and Ks indicated that the studied three species including *Chitala chitala* were undergoing a negative selection, whereas the ND5 and CYTB demonstrated the maximum Ka/Ks values. The phylogenetic tree provided further supplement to the scientific classification of featherback fishes. By using the control region (D-loop) of the mitogenome of the eight studied species, *C. chitala* (GenBank accession no. ON764424) showed 99% similarity of sequence with *C. lopis* (GenBank accession no. AP008922.1), followed by 88% identity with *C. blanci* (GenBank accession no. AP008921.1). The construction of the phylogenetic time tree of *C. chitala* with seven other notopteridae fishes was done using the control regions or D-loop of 8 species from Notopteridae family. The mitogenome of the studied species, *C. chitala* (GenBank accession no.ON764424) displayed divergence time 0.2 million years. The OL region sequence of *C. chitala* predictively had the

ability of folding into a stem-loop stable secondary structure. It comprised the loop of 14 bp and the stem of 11 bp and *Chitala chitala* consumed more free energy than the others and that might be harmful for this species..This study could provide basis information for genetic characters, phylogenetic position and evolution profile for the fishes, which could convenience for resource management or selective breeding in fishery and aquaculture. Further research works should be conducted connecting to the environmental conditions along with the major causes of endangered state of the fish species.

CHAPTER VI
RECOMMENDATIONS

The facts and information presented in this thesis paper may ease more exploration of the molecular evolution and the phylogenetic relationships analysis of Featherbacks. Functions of specific genes should be investigated. The vital molecular resources for the identification of species, management of fishery, and conservation biology regarding Featherbacks were provided by this study. For the protection of the populations of *Chitala chitala*, different plans should be taken into consideration i.e. in-situ conservation by establishing fish sanctuary and protected area for this species, ex-situ conservation by resorting artificial breeding program, stopping, or banning indiscriminate use of insecticides, fertilizers in agricultural fields, practices of bioremediation, treating industrial effluents before mixing to the water bodies for decreasing pollution of their habitats and breeding grounds. At the end, these may be attained by implementing cumulative measures of policy makers, researchers and stakeholders of the country.

REFERENCES

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., & Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature*. **290**(5806): 457–465.

Aguilar, C., Miller, M. J., Loaiza, J. R., Krahe, R., & De León, L. F. (2019). Mitogenomics of Central American weakly-electric fishes. *Gene*. **686**: 164–170.

Anjarsari, Y & Syaifudin, Mochamad & Jubaedah, D & Taqwa, Ferdinand & Yonarta, D. (2021). Phylogenetic of Featherback Chitala sp from South Sumatra Based on Cytochrome C Oxidase Subunit I (COI) Gene. *IOP Conference Series: Earth and Environmental Science*. **8**(10): 009-012.

Asakawa, S., Kumazawa, Y., Araki, T., Himeno, H., Miura, K., & Watanabe, K. (1991). Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes. *Journal of molecular evolution*. **32**(6): 511–520.

Barrientos, A., Barros, M.H., Valnot, I., Rötig, A., Rustin, P., Tzagoloff, A. (2002). Cytochrome oxidase in health and disease. *Gene*. **286**: 53–63.

Banarescu, P. (1990). *Zoogeography of fresh waters*. Aula-Verlag.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. SPAdes: (2012). A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*. **19**: 455–477

Billington N, Hebert PD. (1991). Mitochondrial DNA diversity in fishes and its implications for introductions. *Canadian Journal of Fisheries and Aquatic Sciences*. **48**(S1): 80-94.

Bhuiyan, AL. (1964). Fishes of Dacca. Asiatic Society of Pakistan, Pub. 1, No. 13, Dacca. pp. 15-16.

Boore J. L. (2006). Requirements and standards for organelle genome databases. *Omics : a journal of integrative biology*.**10**(2): 119–126.

Brockman, S. A., & McFadden, C. S. (2012). The mitochondrial genome of *Paraminabea aldersladei* (Cnidaria: Anthozoa: Octocorallia) supports intramolecular recombination as the primary mechanism of gene rearrangement in octocoral mitochondrial genomes. *Genome biology and evolution*. **4**(9): 994–1006.

Catarino, D., Knutsen, H., Veríssimo, A., Olsen, E.M., Jorde, P.E., Menezes, G., Sannæs, H., Stankovi´c, D., Company, J.B., Neat, F. (2015). The Pillars of Hercules as a bathymetric barrier to gene flow promoting isolation in a global deep-sea shark (*Centroscymnus coelolepis*). *Molecular Ecology*. **24**: 6061–6079

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*. **134**(4): 1289–1303.

Chen, T., Qian, J., Dai, Q., Pan, C., Li, B., & Zhang, D. (2021). The complete mitochondrial genome of *Reticulitermes ovatilabrum* (Isoptera: Rhinotermitidae). *Mitochondrial DNA. Part B, Resources.* **6**(3): 794–795.

Chen, X., Xiang, D., Arunrugstichai, S., Cai, L., & Xu, Y. (2016). Complete mitochondrial genome of the mangrove whipray Himantura granulata (Myliobatiformes: Dasyatidae). *Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis*. **27**(1): 1–2.

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

Cheng, J., Ma, G. Q., Song, N., & Gao, T. X. (2012). Complete mitochondrial genome sequence of bighead croaker Collichthys niveatus (Perciformes, Sciaenidae): a mitogenomic perspective on the phylogenetic relationships of Pseudosciaeniae. *Gene*. **491**(2): 210–223.

Chondar, S. L. (1999). Biology of finfish and shellfish. SCSC Publishers (India), West Bengal.

Day, F. (1878). Fishes of India, being a natural history of fishes known to inhabit the seas and freshwaters of India, Burma and Ceylon. William Dawson & Sons Ltd., London, Vol. I: 778 pp.

Desmet, P. G., Cowling, R. M., Ellis, A. G., & Pressey, R. L. (2002). Integrating biosystematic data into conservation planning: perspectives from southern Africa's Succulent Karoo. *Systematic Biology*. **51**(2): 317-330.

Dutta, N., Singh, R. K., Pathak, A., Mohindra, V., Mandal, S., Kaur, G., & Lal, K. K. (2020). Mitochondrial DNA markers reveal genetic connectivity among populations of Osteoglossiform fish Chitala chitala. *Molecular biology reports*. **47**(11): 8579–8592.

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. **32**: 1792–1797

Faith, J. J., & Pollock, D. D. (2003). Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. *Genetics*. **165**(2): 735–745.

Fonseca, M. M., Harris, D. J., & Posada, D. (2014). The inversion of the Control Region in three mitogenomes provides further evidence for an asymmetric model of vertebrate mtDNA replication. PloS One. **9**(9): e106654.

Fuglsang A. (2004). The 'effective number of codons' revisited. *Biochemical and biophysical research communications*. **317**(3): 957–964.

Glémin S. (2003). How are deleterious mutations purged? Drift versus nonrandom mating. Evolution; *International journal of organic evolution*. **57**(12): 2678–2687.

Greenwood, P. H. & Wilson, M. V. (1998). Paxton, J. R. & Eschmeyer, W. N. (eds.). Encyclopedia of Fishes. San Diego: Academic Press. pp. 82–83.

Grant, J.R., Stothard, P. The CGView Server: (2008). A comparative genomics tool for circular genomes. *Nucleic Acids Research*. **36**: 181–184

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

Guan, F., Sheng, Q., Zhang, Y., Lv, H., & Wang, Y. (2022). The complete mitochondrial genome of a vulnerable mandarin fish *Coreoperca liui* (Teleostei: Perciformes: Serranidae) from Qiandaohu Lake in China. *Mitochondrial DNA. Part B, Resources*. **7**(6): 1075–1077.

Gupta, S. K., & Ghosh, T. C. (2001). Gene expressivity is the main factor in dictating the codon usage variation among the genes in Pseudomonas aeruginosa. *Gene*. **273**(1): 63–70.

Gustafsson, C., Govindarajan, S., Minshull, J. (2004). Codon bias and heterologous protein expression. *Trends in Biotechnology*. **22**: 346–353.

Hamilton F., (1822). An account of the fishes found in the river Ganges and its branches, Archibald Constable and Company, Edinburg. 405 pp.

Hassan, M. M., Hussain, M. A., Kambal, S., Elshikh, A. A., Gendeel, O. R., Ahmed, S. A., Altayeb, R. A., Muhajir, A. M., & Mohamed, S. B. (2020). NeoCoV Is Closer to MERS-CoV than SARS-CoV. *Infectious diseases*. **13**: 1178633720930711.

Hayward, B. E., Warner, J. P., Dunlop, N., Fantes, J., Intody, S., Leek, J., Markham, A. F., & Bonthron, D. T. (1997). Molecular genetics of the human glucokinase regulator-fructokinase (GCKR-KHK) region of chromosome 2p23. *Biochemical Society transactions*. **25**(1): 140–145.

He, Z., Gan, H., & Liang, X. (2019). Analysis of Synonymous Codon Usage Bias in Potato Virus M and Its Adaption to Hosts. *Viruses*. **11**(8): 752.

Helfenbein, K. G., Brown, W. M., & Boore, J. L. (2001). The complete mitochondrial genome of the articulate brachiopod Terebratalia transversa. *Molecular biology and evolution*. **18**(9): 1734–1744.

Hilton, Eric J., Lavoué, Sébastien (2018). A review of the systematic biology of fossil and living bony-tongue fishes, Osteoglossomorpha (Actinopterygii: Teleostei). *Neotropical Ichthyology*. **16**(3): 1-35.

Hixson, J.E. & Wong, T.W. & Clayton, D.A.. (1986). Both the Conserved Stem-Loop and Divergent 5'-Flanking Sequences Are Required for Initiation at the Human Mitochondrial Origin of Light-Strand DNA-Replication. *The Journal of biological chemistry*. **261**: 2384-90.

Hogan, R. I., Hopkins, K., Wheeler, A. J., Allcock, A. L., & Yesson, C. (2019). Novel diversity in mitochondrial genomes of deep-sea Pennatulacea (Cnidaria: Anthozoa: Octocorallia). *Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis*. **30**(6): 764–777.

Hong, Y., Duo, H., Hong, J., Yang, J., Liu, S., Yu, L., & Yi, T. (2017). Resequencing and comparison of whole mitochondrial genome to gain insight into the evolutionary status of the Shennongjia golden snub-nosed monkey (SNJ *R. roxellana). Ecology and evolution.* **7**(12): 4456–4464.

Hurst L. D. (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends in genetics : TIG*. **18**(9): 486.

Ikemura T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular biology and evolution*. **2**(1): 13–34.

Imre, Mirela & Văduva, Cristina & Dărăbuș, Gheorghe & Morariu, Sorin & Herman, Viorel & Plutzer, Judit & Suici, Tijana & Lait, Philippa & Imre, Kálmán. (2020). Molecular detection of hemotropic mycoplasmas (hemoplasmas) in domestic cats (Felis catus) in Romania. *BMC Veterinary Research*. **16**: 399.

Inoue, J. G., Kumazawa, Y., Miya, M., & Nishida, M. (2009). The historical biogeography of the freshwater knifefishes using mitogenomic approaches: a mesozoic origin of the Asian notopterids (Actinopterygii: Osteoglossomorpha). *Molecular phylogenetics and evolution*. **51**(3): 486–499.

Islam MN, Sultana S, Alam MJ (2020). Sequencing and annotation of the complete mitochondrial genome of a threatened labeonine fish, Cirrhinus Reba. *Genomics & Informatics*. **18**(3): e32.

Islam, Mohammad Nazrul & Sultana, Shirin. (2022). Codon usage bias and purifying selection identified in Cirrhinus reba mitogenome. *Journal of Advanced Biotechnology and Experimental Therapeutics*. **5**(3): 605-614.

IUCN (2015). Red List of Bangladesh Volume 5: Freshwater Fishes. IUCN, International Union for Conservation of Nature, Bangladesh Country Office, Dhaka, Bangladesh. 2015; 16:360.

IUCN Bangladesh (2015a). Red List of Bangladesh: A Brief on Assessment Result 2015. IUCN, International Union for Conservation of Nature, Bangladesh Country Office, Dhaka, Bangladesh. pp. 24.

IUCN Bangladesh (2015b). Red List of Bangladesh Volume 5: Freshwater Fishes. IUCN, International Union for Conservation of Nature, Bangladesh Country Office, Dhaka, Bangladesh. pp xvi+360

Jia W, Higgs PG (2008). Codon usage in mitochondrial genomes: Distinguishing context-dependent mutation from translational selection. *Molecular Biology and Evolution*. **25**: 339–351.

Jia, C., Zhang, X., Xu, S., Yang, T., Yanagimoto, T., & Gao, T. (2020). Comparative analysis of the complete mitochondrial genomes of three rockfishes (Scorpaeniformes, Sebastiscus) and insights into the phylogenetic relationships of Sebastidae. *Bioscience reports*. **40**(12).

Jin, X., Zhao, S., & Sun, Y. (2013). Complete mitochondrial genome sequence and structure of control region of the Amoya chusanensis (Perciformes, Gobioidei*). Mitochondrial DNA*. **24(**4): 403–405.

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

Junqueira, A. C., Lessinger, A. C., Torres, T. T., da Silva, F. R., Vettore, A. L., Arruda, P., & Azeredo Espin, A. M. (2004). The mitochondrial genome of the blowfly Chrysomya chloropyga (Diptera: Calliphoridae). *Gene*. **339**: 7–15.

Kimura M. (1968). Evolutionary rate at the molecular level. *Nature*. **217**(5129): 624–626.

Korber B. (2000). HIV Signature and Sequence Variation Analysis. In: Rodrigo AG and Learn GH (eds). Computational Analysis of HIV Molecular Sequences. Kluwer Academic Publishers: Dordrecht, Netherlands. pp 55-72.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution*. **33**(7): 1870–1874.

Laslett, D., Canbäck, B. ARWEN: (2007). A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics*. **24**: 172–175.

Liu, Z., Song, N., Yanagimoto, T.,Han, Z., Shui, B., Gao, T. (2017). Complete mitochondrial genome of three fish species (Perciformes: Amblyopinae): Genome description and phylogenetic relationships. *Pakistan Journal of Zoology*. **49**: 107–115.

Liu, Y., Song, F., Jiang, P., Wilson, J.J., Cai, W., Li, H. (2018). Compositional heterogeneity in true bug mitochondrial phylogenomics. *Molecular Phylogenetics and Evolution*. **118**: 135–144.

Loewe, L., & Hill, W. G. (2010). The population genetics of mutations: good, bad and indifferent. Philosophical transactions of the Royal Society of London. Series B, *Biological sciences*. **365**(1544): 1153–1167.

Lowe, T.M., Chan, P.P. (2016). tRNA scan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*. **44**: 54–57.

Lü Z, Zhu K, Jiang H, Lu X, Liu B, Ye Y, Jiang L, Liu L, Gong L. (2019). Complete mitochondrial genome of *Ophichthus brevicaudatus* reveals novel gene order and phylogenetic relationships of Anguilliformes. *International Journal of Biological Macromolecules*. **135**: 609-18.

Lundberg, J. G. (1993). African-South American freshwater fish clades and continental drift: problems with a paradigm. *Biological relationships between Africa and South America*. 156-199.

Lynch, M., Koskella, B., & Schaack, S. (2006). Mutation pressure and the evolution of organelle genomic architecture. *New York Science Journal.* **311**(5768): 1727–1730.

Ma, X. X., Cao, X., Ma, P., Chang, Q. Y., Li, L. J., Zhou, X. K., Zhang, D. R., Li, M. S., & Ma, Z. R. (2018). Comparative genomic analysis for nucleotide, codon, and amino acid usage patterns of mycoplasmas. *Journal of basic microbiology*. **58**(5): 425–439.

Mandal, A., Mohindra, V., Singh, R. K., Punia, P., Singh, A. K., & Lal, K. K. (2012). Mitochondrial DNA variation in natural populations of endangered Indian feather-back fish, Chitala chitala. *Molecular biology reports*. **39**(2): 1765–1775.

Mazumder, Gulshana & Uddin, Dr Arif & Chakraborty, Supriyo. (2015). Prediction of gene expression and codon usage in human parasitic helminths. *Genes & Genomics*. **39**: 10.

Mazumder, Gulshana & Uddin, Dr Arif & Chakraborty, Supriyo. (2020). Analysis of codon usage pattern of mitochondrial ND genes in Platyhelminthes. *Molecular and Biochemical Parasitology*. **238**.

Mitra, Anisa & Mukhopadhyay, P.K.Mukhopadhyay & Homechoudhuri, Sumit. (2018). An Overview of Biology And Culture Potentials of Humped Featherback Chitala chitala (Hamilton, 1822) – A New Candidate for Aquaculture Diversification. *Reviews in Fisheries Science & Aquaculture*. **26**: 1-10.

Montoya, J., Gaines, G. L., & Attardi, G. (1983). The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. *Cell*. **34**(1): 151–159.

Moritz, C. (2002). Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic biology*. **51**(2): 238-254.

Mu X, Liu Y, Lai M. (2015). Characterization of the Macropodus opercularis complete mitochondrial genome and family Channidae taxonomy using Illumina-based de novo transcriptome sequencing. *Gene*. **559**(2): 189-195.

Nei, M. (1972). Genetic distance between populations. *The American Naturalist*. **106**: 283–292.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*. **39**: 197–218.

Ojala, D., Montoya, J., & Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature*. **290**(5806): 470–474.

Perna, N. T., & Kocher, T. D. (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of molecular evolution*. **41**(3): 353–358.

Posada, D. (2008). j ModelTest: Phylogenetic model averaging. Molecular Biology and Evolution. **25**: 1253–1256.

Puppo, P., Curto, M., Meimberg, H. (2016). Genetic structure of Micromeria (Lamiaceae) in Tenerife, the imprint of geological history and hybridization on within-island diversification. *Ecology and Evolution*. **6**: 3443–3460.

Rahman, AKA. (1989). Freshwater Fishes of Bangladesh, 1st edition, *Bangladesh Journal of Zoology*, Zoological Society of Bangladesh, Department of Zoology, University of Dhaka, Dhaka-1000. pp. 237-238.

Ruan, H., Li, M., Li, Z., Huang, J., Chen, W., Sun, J., Liu, L., & Zou, K. (2020). Comparative Analysis of Complete Mitochondrial Genomes of Three *Gerres* Fishes (Perciformes: Gerreidae) and Primary Exploration of Their Evolution History. *International journal of molecular sciences*. **21**(5): 1874.

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

Satoh, T. P., Miya, M., Mabuchi, K., & Nishida, M. (2016). Structure and variation of the mitochondrial genome of fishes. *BMC genomics*. **17**(1): 719.

Schaack, S., Ho, E.K., Macrae, F. (2019). Disentangling the intertwined roles of mutation, selection and drift in the mitochondrial genome. *Philosophical Transactions of the Royal Society B Biological Sciences*. **375**(1790): 20190173

Sharp, P. M., Tuohy, T. M., & Mosurski, K. R. (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic acids research*. **14**(13): 5125–5143.

Sharp, P. M., & Li, W. H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic acids research*. **15**(3): 1281–1295.

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

Shi, X., Tian, P., Lin, R., Huang, D., & Wang, J. (2016). Characterization of the Complete Mitochondrial Genome Sequence of the Globose Head Whiptail Cetonurus globiceps (Gadiformes: Macrouridae) and Its Phylogenetic Analysis. *PloS One*. **11**(4): e0153666.

Sima, Y. H., Chen, M., Yao, R., Li, Y. P., Liu, T., Jin, X., Wang, L. P., Su, J. F., Li, X. S., & Liu, Y. Q. (2013). The complete mitochondrial genome of the Ailanthus silkmoth, Samia cynthia cynthia (Lepidoptera: Saturniidae). *Gene*. **526**(2): 309–317.

Singh, R.K., Dutta, N., Pathak, A. (2019). The complete mitochondrial genome of Indian featherback *Chitala chitala* (Osteoglossiformes) and phylogenetic status. *Conservation Genetics Resources*. **11**: 473–480.

Shen, X., Pu, Z., Chen, X., Murphy, R.W., Shen, Y. (2019). Convergent Evolution of Mitochondrial Genes in Deep-Sea Fishes. *Frontiers in Genetics*. **10**: 925.

Sheng, J. Q., Wang, J. H., He, S. H., Zeng, L. G., Peng, K., & Hong, Y. J. (2014). Complete mitochondrial genome of a natural triploid crucian carp mutant, Carassius auratus var. pingxiangnensis, and phylogenetic analysis of different ploidies in crucian carp. *Genetics and molecular research : GMR*. **13**(3): 5849–5864.

Sui, X., Liang, Y., & He, D. (2016). The complete mitochondrial genome of *Rhynchocypris oxycephalus* (Cypriniformes: Cyprinidae). *Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis*. **27**(5): 3367–3369.

Sweere, T.C., Dickson, A.J., Jenkyns, H.C., Porcelli, D., Elrick, M., Boorn, S.H., Henderson, G.M. (2018). Isotopic evidence for changes in the zinc cycle during Oceanic Anoxic Event 2 (Late Cretaceous). *Geology*. **46**. 463–466.

Talwar, PK. and Jhingran, AG. (1991). Inland Fishes of India and Adjacent Countries, Vol. I, Oxford & IBH Publishing Co. Pvt. Ltd. New Delhi-Calcutta. pp. 63-64.

Tamura K. and Nei M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*. **10**: 512-526

Tamura K., Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, and Kumar S. (2012). Estimating Divergence Times in Large Molecular Phylogenies. *Proceedings of the National Academy of Sciences*. **109**:19333-19338

Tamura K., Stecher G., and Kumar S. (2021). MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution.* **10**. 115-120.

Tatarinova, T.V., Alexandrov, N.N., Bouck, J.B.; Feldmann, K.A. (2010). GC 3 biology in corn, rice, sorghum and other grasses. *BMC Genomics*. **11**: 308.

Uddin, A., Choudhury, M. N., & Chakraborty, S. (2016). Codon usage bias and phylogenetic analysis of mitochondrial ND1 gene in pisces, aves, and mammals. *Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis.* **29**(1): 36–48.

Urbina, D., Tang, B., & Higgs, P. G. (2006). The response of amino acid frequencies to directional mutation pressure in mitochondrial genome sequences is related to the physical properties of the amino acids and to the structure of the genetic code. *Journal of molecular evolution*. **62**(3): 340–361.

Wang, C., Ye, P., Liu, M., Zhang, Y., Feng, H., Liu, J., Zhou, H., Wang, J., & Chen, X. (2022). Comparative Analysis of Four Complete Mitochondrial Genomes of Epinephelidae (Perciformes). *Genes*. **13**(4): 660.

Waldbieser, Geoffrey & Bourgeois, Lanie & Nonneman, Dan. (2003). Complete Sequence and Characterization of the Channel Catfish Mitochondrial Genome. *DNA sequence : the journal of DNA sequencing and mapping*. **14**: 265-77.

Wei, S.J., Shi, M., Sharkey, M.J., van Achterberg, C., Chen, X.X. (2010). Comparative mitogenomics of Braconidae (Insecta: Hymenoptera) and the phylogenetic utility of mitochondrial genomes with special reference to Holometabolous insects. *BMC Genomics*. **11**: 371

Weinstein JN, Myers TG, O'Connor PM (1997). An information intensive approach to the molecular pharmacology of cancer. Science. **275**: 343–349.

Wibowo, Arif. (2010). Genetic Differentiation of the Kampar River's Giant Featherback (Chitala lopis Bleeker 1851) Base on Mitochondrial DNA Analysis. Indonesian Fisheries Research Journal. **16**: 49-58.

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

Wu, Y., Zhao, D., & Tao, J. (2015) Analysis of Codon Usage Patterns in Herbaceous Peony (Paeonia lactiflora Pall.) Based on Transcriptome Data. *Genes*. **6(**4): 1125–1139.

Wu, H.L., Shao, G.Z., Lai, C.F., Zhuang, L.H., Lin, P.L. (2017*)*. Latin-Chinese Dictionary of Fish Names by Classification System; China Ocean University Press: Qingdao, China. p. 602.

Xiang, D., Shen, X., Pu, Z., Irwin, D.M., Liao, M., Shen, Y. (2018). Convergent evolution of human-isolated H7N9 avian influenza A viruses. Journal of Infectious Diseases. **217**: 1699–1707.

Yan, Y., Yao, D., Tian, Z.X., Huang, C.Y., Dilek, Y., Clift, P.D., Li, Z.A. (2018) Tectonic topography changes in Cenozoic East Asia: A landscape erosion-sediment archive in the South China Sea. *Geochemistry, Geophysics, Geosystems*. **19**: 1731–1750.

Yang, Z., Bielawski, J.P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecological Evolution*. **15**: 496–503.

Yang, H., Xia, J., Zhang, J. E., Yang, J., Zhao, H., Wang, Q., Sun, J., Xue, H., Wu, Y., Chen, J., Huang, J., & Liu, L. (2018) Characterization of the Complete Mitochondrial Genome Sequences of Three Croakers (Perciformes, Sciaenidae) and Novel Insights into the Phylogenetics. *International journal of molecular sciences*. **19**(6): 1741.

Yengkhom, Sophiarani & Uddin, Dr Arif & Chakraborty, Supriyo. (2019). Deciphering codon usage patterns and evolutionary forces in chloroplast genes of Camellia sinensis var. assamica and Camellia sinensis var. sinensis in comparison to Camellia pubicosta. *Journal of Integrative Agriculture*. **12**: 2771–2785.

Yi, M. R., Hsu, K. C., Gu, S., He, X. B., Luo, Z. S., Lin, H. D., & Yan, Y. R. (2022). Complete mitogenomes of four *Trichiurus* species: A taxonomic review of the *T.lepturus* species complex. *ZooKeys*. **1084**: 1–26.

Yu, J. N., & Kwak, M. (2015). The complete mitochondrial genome of Brachymystax lenok tsinlingensis (Salmoninae, Salmonidae) and its intraspecific variation. *Gene*. **573**(2): 246–253.

Zhang, Z., & Yu, J. (2006). Evaluation of six methods for estimating synonymous and nonsynonymous substitution rates. *Genomics, proteomics & bioinformatics*. **4**(3): 173–181.

Zhu, K.C., Liang, Y.Y., Wu, N., Guo, H.Y., Zhang, N., Jiang, S.G., Zhang, D.C. (2017). Sequencing and characterization of the complete mitochondrial genome of *Japanese Swellshark* (*Cephalloscyllium umbratile*). *Scientific Reports*. **7**: 15299.

Zhao, W., Zhao, Q., Li, M., Wei, J., Zhang, X., Zhang, H. (2019). Comparative Mitogenomic Analysis of the *Eurydema* Genus in the Context of Representative Pentatomidae (Hemiptera: Heteroptera), *Journal of Insect Science*. **19**: 20.

Zuker, M., (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*. **31**: 3406–3415.

APPENDICES

Appendix Table 1. Base composition of 13 PCGs of *Chitala chitala*

PCGs	A%	G%	C%	T%	A+T%	Total (bp)
ND1	29.7	14.6	31.9	23.7	53.4	969
ND2	34.9	10.9	33.2	20.9	55.8	1047
COX1	28.3	16.8	25.9	28.9	57.2	1549
COX2	31.7	15.8	26.6	25.9	57.6	691
ATP8	35.1	7.7	31.5	25.6	60.7	168
ATP6	30.4	11.8	29.7	28.1	58.5	684
COX3	27.9	16.4	29.9	25.7	53.6	786
ND3	27.8	14.3	30.9	26.9	54.7	349
ND4L	25.3	15.5	30.6	28.6	53.9	297
ND4	31.1	13.6	28.7	26.6	57.7	1381
ND5	33.1	11.4	30.4	25.2	58.3	1838
ND6	43.8	12.8	29.8	13.6	57.4	523
CYTB	30.3	13.7	29.3	26.7	57.0	1141
All	409.3	175.4	388.7	326.6	735.9	11423