DNA Barcoding of Marine Clupeid Fishes

(Order- Clupeiformes) of Bangladesh

By
Afroza Kaonine Haque
15136030

A thesis submitted to the Department of Mathematics and Natural Sciences in partial fulfillment of the requirements for the degree of
Bachelor of Science in Biotechnology

Department of Mathematics and Natural Sciences
Brac University
December 2019

Declaration

It is hereby declared that:

1. The thesis submitted is my own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

Afroza Kaonine Haque

15136030

Approval

The thesis titled "DNA Barcoding of marine Clupeid fishes (Order- Clupeiformes) of Bangladesh" submitted by

1. Afroza Kaonine Haque (15136030)

of Spring, 2015 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Science in Biotechnology on 5th December 2019.

Examining Committee:

Supervisor 1:
(Member)

Ms. Romana Siddique
Senior Lecturer, Mathematics and Natural Sciences
Brac University

Supervisor 2:
(Member)

Dr. Md. Sagir Ahmed
Professor, Department of Zoology
University of Dhaka

Program Coordinator:
(Member)

Iftekhar Bin Naser
Assistant Professor, Mathematics and Natural Sciences
Brac University

Departmental Head:
(Chair)

Prof. A. F. M Yusuf Haider
Chairperson, Mathematics and Natural Sciences
Brac University

Ethics Statement

This study has been conducted with samples from the DNA Barcoding Lab, Department of Zoology, Dhaka University and the consent taken from Dr. Md. Sagir Ahmed [Proffessor, Department of Zoology, Dhaka University] in agreement to use the samples for thesis purpose only. The sequence data is strictly confidential, and any further utilization or reproduce of data need prior permission.

ABSTRACT

Objective: To determine the efficacy of DNA barcoding for molecular identification of marine Clupeid fishes (Order- Clupeiformes) using the Cytochrome C Oxidase marker from their mitochondrial DNA and establish the phylogenetic relationship among the Clupeid fishes.

Methods: 80 different samples of marine Clupeid fishes (Order- Clupeiformes) were collected and their DNA extracted and purified, from which 35 different sequences of cytochrome oxidase C was sequenced and identified. Then, homology test was undergone using BLAST and the sequences submitted to GenBank. At last. similar sequences were retrieved from GenBank database and subjected to different bioinformatics tools for phylogenetic analysis.

Results: A total of 185 sequences were studied. These belonged to three families (Clupeidae, Engraulidae and Pristigasteridae), eleven genus and eighteen species. The GC% showed a minimum, maximum and average value of 44.36, 49.09 and 46.14 respectively. The GC% content in all families followed a trend of 1_{st} codon> 2_{nd} codon> 3_{rd} codon positions. On the other hand, the Kimura 2 Parameter (K2P) distance between Intra and Inter species showed a perfect 3 units of difference, which increased when the distances were calculated for Genus and Families. The intraspecies (minimum: 0.26, maximum: 6, average: 1.88) intragenus (minimum: 0, maximum: 12, average: 8.33) and intrafamily distances (minimum: 13, maximum: 20, average: 17) showed an expected gradual increase, demonstrating that the level of genetic variation soars as the species drifts far apart from each other. Also, the phylogenic trees constructed proved the efficacy of DNA barcoding in differentiating species of separate families.

This study demonstrated the applicability of mitochondrial COI gene-based DNA barcoding as a powerful tool to identify, discriminate and classify different fish species, as the results proved great efficacy without any anomalies, and hence make identification possible. In further studies, extended analysis of the marker can help to determine differences caused by geographic location and polymorphic change.

Keywords: DNA Barcoding, Kimura-2-Parameter, GC%, Phylogenetic analysis

Dedication

Dedicated to Abbu, Ammu, and some of my you-know-whos in life who have helped me grow.

With you-know-what, and you certainly know why.

Acknowledgement

I am immensely grateful to Dr. Md. Sagir Ahmed [Professor, Department of Zoology, University of Dhaka] for letting me work under his supervision at his DNA Barcoding Lab, Department of Zoology, University of Dhaka, Ms. Romana Siddique [Senior Lecturer, Mathematics and Natural Sciences Department, Brac University] for all her constant support and encouragement, and Md. Zarif Hossain [Lecturer, Department of Oceanography, University of Dhaka] for enlightening me with his wisdom.

I would also like to acknowledge the support provided by Dr. Md. Sagir Ahmed and his team in the collection and confirmation of Clupeiformes species caught in Bangladeshi waters, as well as letting me work on the confidential data that I analyzed on.

Table of Contents

List of Tables

List of Figures:

List of Acronyms

BLAST: Basic Local Alignment Search Tool

COI: Cytochrome oxidase subunit I

DNA: Deoxyribonucleic acid

K2P: Kimura 2 Parameter

MEGA: Molecular Evolutionary Genetic Analysis

MT COI: Mitochondrially encoded cytochrome C oxidase

mtDNA: mitochondrial DNA

MUSCLE: Multiple Sequence Comparison by Log-Expectation

NCBI: National Center for Biotechnology Information

nDNA: nuclear DNA

PCR: Polymerase Chain Reaction

RNA: Ribonucleic acid

Chapter 1

Introduction

1.1 Background of the study

Fish are the largest group of vertebrates, which exhibit a remarkable diversity of morphological attributes and biological adaptations. Species are typically limited to being identified by the presence of fixed diagnostic morphological characters which distinguish them from other species. But for fishes, there are a large number of intraspecific invariants or interspecific overlapping, so fish identification is challenging for taxonomists when facing rich biota.

The main focus of this study has been the three families sharing species of Ilisha, the most popular fish in Bangladesh. More importantly, Clupeidae is the most valuable family of food fishes in the world. Other examples include Hilsa shad, Indian oil sardine, round sardinella, goldstrip sardinella, chapila, etc. Bangladesh has an annual production of *Tenulosa ilisha* of 12.09% among all other fishes in inland and marine fisheries combined (FRSS, 2018), with an estimated annual production of 517,198 tons, which was 1.0% to the GDP to support livelihoods of 1.2 million hilsa fishers of Bangladesh. Naturally, it is not only celebrated in the country but has high value in international market as well, which is proved by the fact that Bangladesh harvest about 60% of the world hilsa catch. Hence, making sure the quality and taste is top-level throughout time is important. As increase in international trade and global Ilisha consumption, along with fluctuations in the supply and demand of different fish species have resulted in intentional product mislabeling, keeping the grade of the product intact has been alarmingly tough since the past few years. The effects of species substitution are far-reaching and include economic fraud, health hazards, and illegal trade of protected species.

Since two hundred years back when it all started with Aristotle and Linnaeus, classification of species have been inevitably fundamental to research in biodiversity, ecology, evolutionary biology and conservation biology. As a part of this background in the present study, 18 species of fishes from three families: Engraulidae, Clupeidae and Pristigasteridae, were sequenced for their 650bp region of cytochrome oxidase subunit I

(COI) gene, while the rests were extracted from NCBI GenBank for intra and inter species variation analysis. This was undergone to test the efficacy in identifying the species and also to demonstrate the intra species variations within the barcode region.

To improve detection of commercial exchange fraud, a variety of DNA-based techniques have been developed, including Multiplex PCR, FINS, PCR-RFLP, PCR-RAPD, PCR-AFLP, and PCR-SSCP, based on the discovery of polymorphism in protein or deoxyribonucleic acid (DNA) characteristics that are unique to each species (Rasmussen & Morrissey, 2008[1]). Here, the analytical techniques used to establish the unique fingerprint are first to be optimized for the specific product under investigation, only after which they will be able to provide undeniable and repeatable results that prove species identification (Woolfe & Primrose, 2004 [2]). Undoubtedly, complications arise when an individual from the same species show different fingerprints or a number of species have similar fingerprints due to intraspecies variation. In addition, certain processing steps are known to denature proteins and partially degrade DNA, making analysis of the fishes in the market especially demanding.

Indeed, if this identification and discrimination of species takes place on the basis of morphological characteristics, it will require a considerable amount of time, specialized expertise as well as a rigorous effort. Approximately 1.5 million species have been recorded so far and only 0.01 percent has been established by traditional taxonomy to date. Effective inspection and tenacious research of 15,000 devoted taxonomists would entail constant classification of all these animals. As a consequence, DNA-based classification and detection are necessary to address these limitations and therefore the first approaches have started with the smallest species, such as viruses, bacteria, various protists and some fungi. After that, the merit, efficiency, low cost and speed of this technique were addressed and quickly improvised and implemented for higher organisms as well. Areas widely used for DNA barcoding include nuclear DNA (e.g. ITS), chloroplast DNA (e.g. rbcL, trnL-F, matK, psbA, trnH, psbK) and mitochondrial DNA (e.g. COI). Here, the use of this tool increased exponentially after the use of mitochondrial cytochrome c oxidase I (COI) for the classification of different animal species.

DNA Barcoding has some distinctive advantages which makes it a remarkable distinguishing process: Firstly, The use of DNA-based methods for species detection presents a number of advantages over protein-based methods, including increased specificity, sensitivity, and reliable performance with highly processed samples (Lenstra 2003[3]). Although DNA molecules can degrade during processing, they are more thermostable than proteins: DNA fragments as long as 300 bp can still be recovered following sterilization (Chapela, et al., 2007[4]). Also, DNA has the potential to provide a greater amount of information due to the degeneracy of the genetic code and the existence of noncoding regions (Lockley & Bardsley, 2000[5]). Whereas proteins vary with tissue type, age, and status, DNA is largely independent of these factors and is present in all cell types (Bossier, 1999[6]; Civera, 2003[7]). Since analytical methods based on DNA have been shown to have several advantages over those based on proteins, the use of mitochondrial DNA for the process is highly reliable and convenient.

Development of reliable tool for unambiguous discrimination studies of diverse species populated in this planet is not an easy task. Scientists apply DNA barcoding technique by utilizing an exclusive and ultimate source from genetic structure available in the organisms and provide promising and reproducible results with certainty. It is one of the latest concepts, aiming to afford rapid, accurate and automatable species identification technique using a standardized DNA region as a tag. As Chase et al. described, there are two categories of potential DNA barcode users: i) taxonomists and ii) biotechnologists and scientists working in the area of forensic science, food industry, animal diet etc. The recent trend demands the ideal DNA barcoding system to meet these criteria a) It should be sufficiently variable to discriminate among all species, but conserved enough to be less variable within than between species b) It should be standardized, with the same DNA region as far as possible used for various taxonomic groups c) It should be extremely robust, with highly conserved priming site and highly reliable DNA amplifications and sequencing. Unfortunately, such an ideal and comprehensive DNA marker is found to be limited for wide variety of plants. However, the criteria's listed above will not be equally important for different category of users i.e. taxonomists, biodiversity analysts and scientists working in the field of advanced biological sciences because a high level of variation with sufficient phylogenetic information and unique DNA regions are essential. Hence, DNA barcoding proves to be means when it is actually a tool to be used largely for discrimination and identification purposes.

About the MT-CO1 gene sequence used here, it is suitable for its role because it has a mutation rate fast enough to distinguish closely related species, coupling with the fact that its sequence is conserved among conspecifics. Contrary to the primary objection raised by skeptics that MT-CO1 sequence differences are too small to be detected between closely related species, more than 2% sequence divergence is typically detected between closely related animal species (Hebert, et al., 2003[8]), suggesting that the barcode is effective for many animals.

1.2 About DNA Barcoding

Freshwater fishes show more population differentiation than marine species, although marine species can show significant differentiation. Indeed, several studies have already illustrated the advances provided by the iterative processes between morphological- and DNA barcode-based studies in fishes. DNA barcode, a short section of DNA sequence is used to identify species. Neither the idea nor the technology behind DNA Barcoding is novel. What is new and controversial is the idea of using just a small portion of a single gene to identify species from a wide taxonomic range. Hebert et al introduced the concept of a DNA barcode, and proposed a new loom to species identification which offered greater promise to counter many of the limitations. The new approach is based on the ground that the sequence analysis of a short fragment of a single gene “Cytochrome C oxidase subunit 1” enables unambiguous identification of all animals species. Hebert et al suggested a 650 base pair sequence of mitochondrial gene Cytochrome C oxidase subunit 1 (COI) as the reference DNA barcode for all animal life. This gene occurs in the mitochondria of all eukaryotic organisms and the initial studies revealed consistent resolving capability at the species level for many animals. DNA barcoding, which was advocated by Hebert et al seeks to facilitate identifying the increasing number of unfamiliar taxa in biological conservation and biodiversity surveys, based on sequence diversity within a short and standardized gene region. For coordinating the collection data of specimens and performing data analysis with barcode data, the GenBank of NBCI has been used.

Animal mtDNA contains 1 major non- coding region, 13 protein-coding genes, 22 genes coding for transfer ribonucleic acid (tRNA), and 2 genes coding for ribo- somal RNA (rRNA) (Cespedes, et al., 2000[9]). Some major advantages of mtDNA over nDNA are

that it is relatively simple and small compared to nDNA because it lacks features such as large noncoding sequences (introns), pseudogenes, repetitive DNA, and transposable elements; it is relatively easy to extract; it does not undergo genetic rearrangements such as recombination; and sequence ambiguities resulting from heterozygous genotypes are avoided (Cespedes, et al., 2000[9]; Civera, 2003[7] ; Aranishi, et al., 2005[10]). In addition, mtDNA- which is maternally inherited- exhibits a higher copy number and a faster rate of mutation, making it generally more appropriate in the study of evolutionary genetics and inter and intraspecies variability (Carrera, et al., 2000b (48)). Due to the widespread use of mtDNA in genetic research, many universal primers have already been designed to facilitate the amplification of mtDNA fragments for fish and seafood species diagnostics (Carrera, et al., 2000a[11]; Comesana, et al., 2003[12]).

1.3 Objectives of DNA Barcoding

The goal of DNA barcoding is conceptually simple, and it urges to find one or a few regions of DNA which facilitates distinguishing among the majority of the world's species, and sequence these from diverse sample sets to produce a macroscopic reference library of existence on earth. By taking into consideration the well-established molecular biology techniques and emerging developments in bioinformatics, DNA barcoding offers the opportunity to make use of biodiversity studies in entirely a new way. DNA barcoding methods have wide-spread applications to help protect biodiversity against such threats as man-made changes in the environment and the pervasive illegal commercial trade in animals and their products. First DNA barcoding studies were made in animal and a portion of the mitochondrial gene Cytochrome Oxidase 1 which has proved remarkably effective at discriminating among species in diverse groups such as birds, fishes, and insects. The identification of animal biological diversity by using molecular markers has recently been proposed and demonstrated on a large scale through the use of a short DNA sequence in the cytochrome c oxidase 1 (CO1) gene. These ''DNA barcodes'' are promising in case of providing a practical, standardized, species-level identification tool that proves to be helping hand for biodiversity assessment, life history, ecological studies, and forensic analysis. Engineered DNA sequences also have been suggested as exact identifiers and intellectual property tags for transgenic organisms. Tracing any one source of particular organism for developing DNA barcode will not solve the purpose, for example chloroplast organelle sequence for plants or

mitochondrial DNA sequences for animals may not be sufficient for the entire group of organisms.

1.4 DNA Barcoding Complementing Conventional Taxonomy

Conventional taxonomy requires a lot of tenacious effort as well as a long period of time. Herbert addresses 4 specific limitations of taxonomy based on morphological characteristics of a species:

-Incorrect conclusions can be drawn because of genetic variability as well as phenotypic plasticity in case of recognizing different species.

-Secondly, conventional taxonomy overlooks morphologically cryptic taxa that are found to be common in many groups.

-Thirdly, morphological keys remain effective only for a particular stage in the life cycle, certain metamorph, and a particular gender. As a result, many individuals remain unidentified.

-Finally, the understanding and proper implementation of these keys require high level of expertise, tenacious effort and perpetual amount of time.

But morphological identification still stands as a gold standard of taxonomical classification partly also, due to some limitations of DNA barcoding. Nevertheless, since the ease of this technology as well as cost effectiveness and attractive swiftness, it has have made itself as an inexpendable approach to modern taxonomy and molecular phylogeny.

1.5 Locus Used for DNA Barcoding

Up until 2003, different laboratories used different DNA stretches for barcoding organisms. It posed difficulty because comparing the results become impossible without a common standard. So, a necessity of choosing a standard locus seemed necessary to compare the results derived from different laboratories. A certain locus for DNA barcoding is chosen. Thus, creation of a common and standard yardstick to compare in between species and furthermore constructing a database becomes possible. This requires certain qualities such as:

-Presented in almost most of the taxa so that PCR amplification can be performed using a specific primer.
-Short stretch of DNA from which the amplicon can be conveniently sequenced with the available technology.
-Provides large interspecies variations, but limited intraspecies variations, making discrimination between species possible.

But it was not possible to establish such a common locus among all living organisms. Still, for different supergroups of species, different loci have been recorded so far to differentiate between individuals. They are:
-Mitochondrial COI gene for animals and certain eukaryotes
-Chloroplast rbcL and matK for plants
-Internal transcribed spacer (ITS) region for fungi

1.6 Cytochrome C Oxidase I (COI*)*

Up until 2003, different laboratories used different DNA stretches barcode. So a common platform was necessary which was resolved by the presence of mitochondria in all of the eukaryotes. Animal mitochondrial DNA (mtDNA) is characterized with comparatively higher mutation rate. As a result, over a short period of evolutionary timescale, observable variations in between and within populations occur. Additionally, animal mitochondria is inherited from mother and for these two reasons, larger divergence of mtDNA sequences between species occur whereas comparatively small variance within species is observed. The 655 bp region of the mitochondrial cytochrome c oxidase subunit I (COI or COX1) gene, namely the Folmer region is associated with highest variation and thus considered as the 'barcode' of animals and some eukaryotes. Cytochrome c oxidase or respiratory complex IV spans across the membrane (transmembrane protein) of eukaryotes. It is a key enzyme in the mitochondrial oxidative phosphorylation process and works as the final enzyme of the electron transport chain. Its mutation rate is high to discriminate different species whereas its sequence is conserved in the individuals belonging to the same species.

Mitochondrial genes are preferred over nuclear genes because of their lack of introns, their haploid mode of inheritance and their limited recombination. Moreover,

each cell has various mitochondria (up to several thousand) and each of them contains several circular DNA molecules. Mitochondria can therefore offer abundant source of DNA even when sample tissue is limited.

Chapter 2

Materials and Method

2.1 Study species

Clupeiformes is the order of ray-finned fish that includes the herring family, Clupeidae, the long fin herring family Pristigasteridae and the anchovy family, Engraulidae. The group includes many of the most important forage and food fish. Clupeiformes are physostomes, which means that the gas bladder has a pneumatic duct connecting it to the gut. They typically lack a lateral line, but still have the eyes, fins and scales that are common to most fish. They are generally silvery, with streamlined, spindle-shaped bodies, and they often school. They have a single short-based dorsal fin, abdominal pelvic fins and a forked tail. Most species ingest plankton which they filter from the water with their gill rakers.

Among five of its families, three families of fishes which were available in Bangladesh's region were selected and worked on. They are:

Family- Clupeidae

A family of ray-finned fishes, comprising, for instance, the herrings, shads, sardines, hilsa, and menhadens. These include many of the most important food fishes in the world and are also commonly caught for production of fish oil and fish meal. Many members of the family have a body protected with shiny cycloid- very smooth and uniform- scales, a single dorsal fin, and a fusiform body for quick, evasive swimming.

Morphological detection:

Clupeids are mostly marine forage fish, although a few species are found in fresh water. No species has scales on the head, and some are entirely scale less. The lateral line is

short or absent, and the teeth are unusually small where they are present at all. Clupeids typically feed on plankton, and range from 2 to 75 cm (0.8 to 30 in.) in length.

Family- Engraulidae

Most species are found in marine waters, but several will enter brackish water and some in South America are restricted to fresh water.

Anchovies are small, green fish with blue reflections due to a silver-colored longitudinal stripe that runs from the base of the caudal (tail) fin. They range from 2 to 40 cm (0.79 to 15.75 in) in adult length, and their body shapes are variable with more slender fish in northern populations. The snout is blunt with tiny, sharp teeth in both jaws. The snout contains a unique rostral organ, believed to be sensory in nature, although its exact function is unknown. The mouth is larger than that of herrings and silversides, two fish which anchovies closely resemble in other respects. The anchovy eats plankton and recently hatched fish.

Family- Pristigasteridae

Distribution: Atlantic, Pacific, and Indian oceans. Tropical oceans and in the freshwaters of South America and Southeast Asia. Externally distinguished from other sardines (Clupeidae) by their long anal fin, which has at least 30 fin rays; in nearly all members, the small pelvic fin is markedly displaced anteriorly that the tip of the pectoral fin reaches or surpasses the vertical through the base of the pelvic fin.

Figure 1. Some common commercially important Clupeid fishes: A. *Stolephorus dubiosus,* B. *Coilia dussumieri,* C. *Pellona ditchela.*

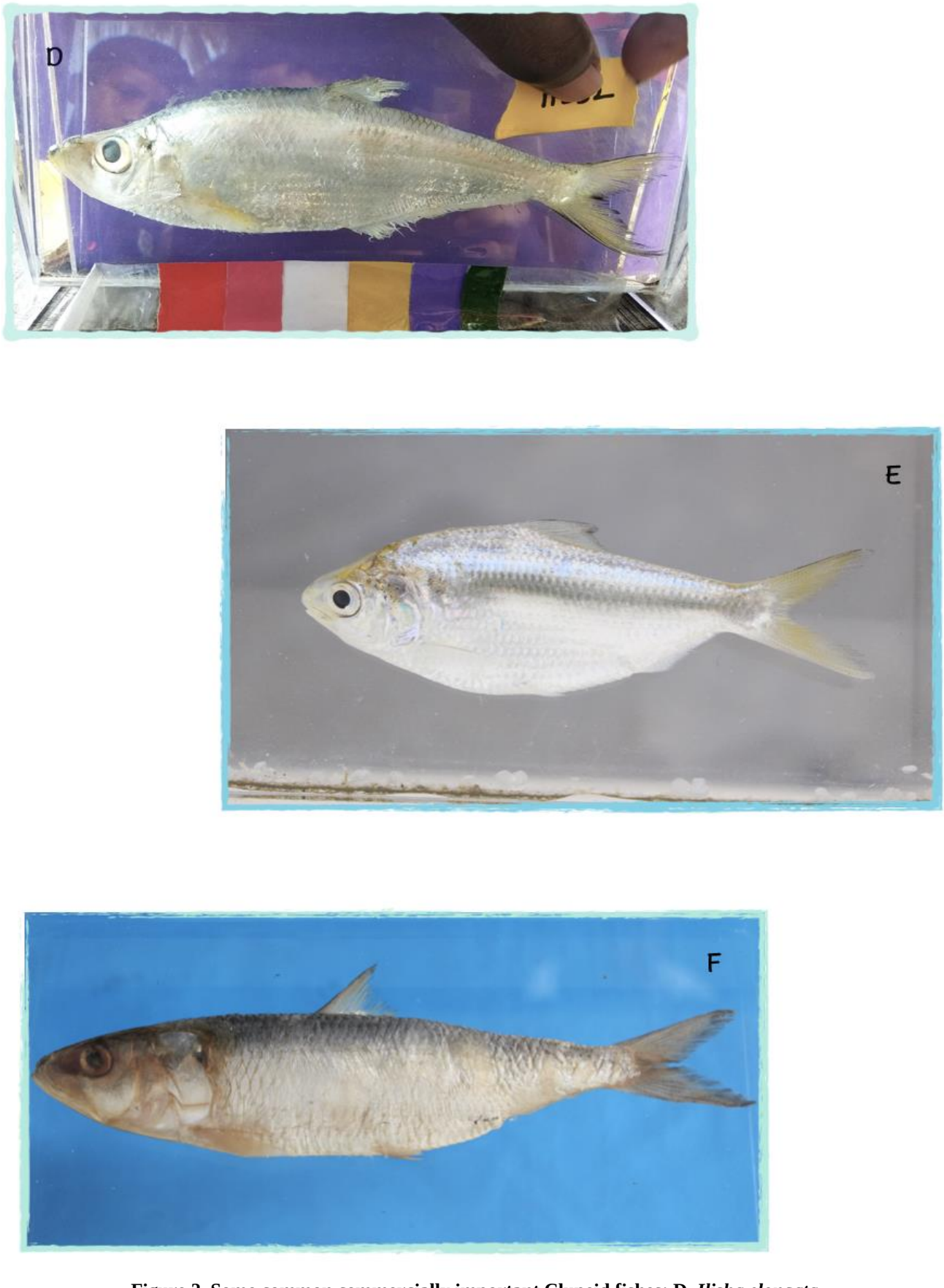

Figure 2. Some common commercially important Clupeid fishes: D. *Ilisha elongata,*

E. *Anodontostoma chacunda*, F. *Sardinella longiceps.*

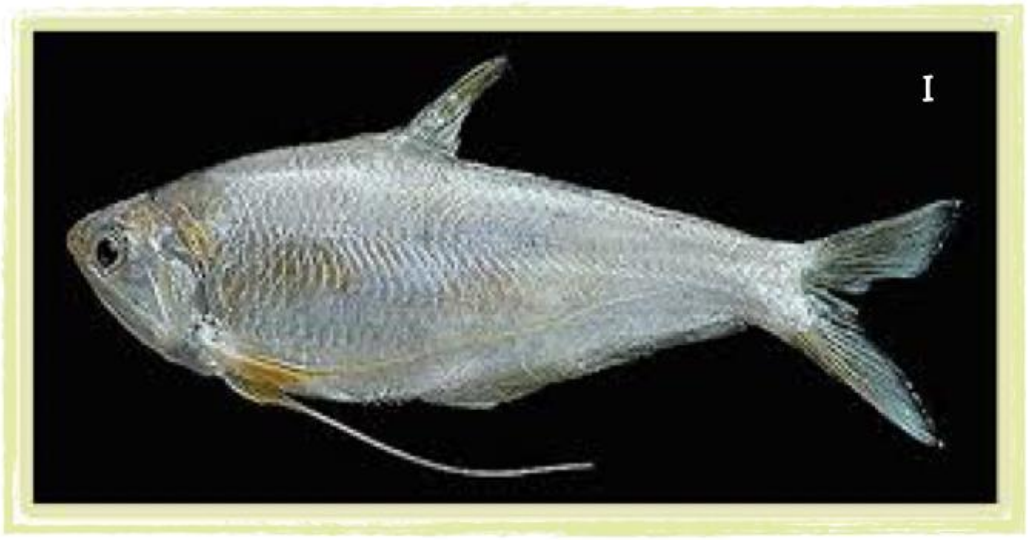

Figure 3. Some common commercially important Clupeid fishes: G. *Opisthopterus tardoore*, H. *Tenualosa ilisha*, I. *Setapphinna phasa.*

2.2 DNA extraction

Genomic DNA was extracted from the stored muscle tissue samples by the standard Proteinase- K/Phenol–Chloroform– ethanol method. The concentration of DNA was estimated using a UV spectrophotometer. The COI gene located in the mitochondrial genome was amplified using two sets of primers and sent to Malaysia for Sanger sequencing.

1. Equal volume of Phenol: Chloroform: Isoamylalcohol (25:24:1) was added and mixed thoroughly in an up-and-down fashion for few minutes.

2. The samples were centrifuged at 12000 rpm for 15 minutes.

3. Upper aqueous phase was transferred to a new Eppendorf tube and equal volume of Chloroform: Isoamylalcohol was added. Then it was centrifuged at 12000 rpm for 15 minutes.

4. Upper aqueous phase was transferred to a new Eppendorf tube. Double volume of chilled absolute ethanol (100%) was added to it.

5. The samples were again centrifuged at 12000 rpm for 15 minutes. After this step DNA precipitated as pellet. The supernatant was discarded.

6. To wash the DNA 70% ethanol (v/v) was added. They were again centrifuged at 12000 rpm for 15 minutes. Pellet was procured discarding the supernatant and furthermore air-dried.

7. The pellet was resuspended in Nuclease Free Water (NFW) stored in -20° C for long preservation.

2.3 Gel Electrophoresis and Observation of DNA Bands

1. DNA was mixed with the loading dye bromophenol blue and loaded into 1% (w/v) agarose gel. The gel was prepared using agarose powder and 1X TAE buffer with ethydium bromide being dissolved.

2. The DNA was electrophoresed at 110 V for 30 minutes in 1X TAE buffer. The bands were separated thus and furthermore observed. The marker used for DNA ladder 1 kb.

3. The gel was observed in the gel documenter, AlphaImager HP using UV light.

4. Image was retrieved in the computer using AlphaImager HP software. The contrast and the background was manually set to get the best of the views. Whether DNA was extracted in the first hand was decided by comparing the respective bands with a fixed corresponding ladder.

2.4 PCR Amplification

1. The extracted DNA samples were used to perform molecular assessment via Polymerase Chain Reaction (PCR) to observe the presence of particular genes. PCR is a key technique in molecular genetics that permits the analysis of any short sequence of DNA through amplification of a single copy or a few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence.

2. The major assay currently used in different fish species identification is based on PCR amplification which requires much less starting material (2µL of DNA) and exhibits greater versatility and sensitivity.

3. Amplification of genetic material with PCR was done with Taq polymerase, 2 oligonucleotide primers FishF1 and FishR1 (Wards, et al., 2005[17]), 4 deoxynucleotide triphosphates (dNTPs). The final volume was 25µL.

4. The PCR involved 3 reaction steps carried out at different temperatures: denaturation (approximately 95 °C), annealing (50 to 60 °C), and extension (approximately 72 °C). During these 3 steps, the template DNA was first separated into 2 single strands by heat denaturation, then the oligonucleotide primers annealed to complementary sequences on opposing ends of a particular fragment of the template DNA, and lastly

Taq polymerase utilized the 4 dNTPs to synthesize multiple copies of the target DNA fragment.

5. 50 cycles of these three steps namely: denaturation, annealing, and extension were performed in the PCR machine.

6. For agarose gel electrophoresis 10µl of DNA products from PCR amplification was loaded in 1.2% agarose gel in 1X TAE buffer. Then after running 30 min, it was observed under UV transilluminator. Furthermore, it was documented with Gel-DOC.

The primer sequences and the constituents of the master mix are described in the tables below.

Table 1. Description of the primers used for COI gene amplification

Forward primer (F1)	Reverse primer (R1)
5'TCAACCAACCACAAAGACATTGGCAC3' [5]	3'TAGACTTCTGGGTGGCCAAAGAATCA5' [5]
26 Bases	26 Bases
Tm: 58°C	Tm: 58°C

Table 2. Composition of the Master Mix

Components	Volume (µl)
Taq polymerase	**12.5**
Forward primer (F1)	**1**
Reverse primer (R1)	**1**
Extracted DNA	**2**
Nuclease free water	**8.5**
Total	**25**

2.5 Purification of PCR product

1. Same amount of loading buffer was added in the PCR tubes (25µL of loading buffer in case of 25µL of PCR being produced).

2. The tubes were then put into the spinner mix for 15-20 seconds.

3. Next, the whole amount was transferred into the purification tubes.

4. The tubes were now microcentrifuged for 1 minute, at 12,000 rpm.

5. After discarding the tubes with the elutes in them, the suckers were taken into fresh, pre-sterilized eppendorf tubes.

6. 700µL wash buffer was pipetted into each of the sucker.

7. The tubes were centrifuged for 1 min at 12,000-14,000 rpm.

8. After discarding the liquid from the columns, the tubes were centrifuged for 1 min at 12,000-14,000 rpm for the second time.

9. While the used columns were discarded, new ones were filled with 40µL of elution buffer in each of them.

12. For the last time, the tubes were then centrifuged for 1 min at 12,000 rpm.

13. Discarding the sucker, the columns which now contained purified PCR product were preserved at -20° C.
14. Later, the purified PCR products were loaded in the 1.2% (w/v) agarose gel and electrophoresed at 110 V for 30 minutes, followed by observing under gel documenter, eventually producing an electropherogram.

2.6 DNA Quantification and Sequencing

1. Quantitation of DNA was estimated in three different absorbances and the average was taken into consideration. Nano drop spectrophotometer was used for this purpose.

2. After that, the purified PCR products were sent to Malaysia for sequencing.

2.7 Using Chromas

Chromas is a free, simple, easy-to-use viewer and editor for chromatograms (traces) from automated Sanger sequencers. It has many format conversion options including batch processing functions to handle many files at once. As the mitochondrial COI gene sequence of respective species were delivered as chromatogram file format (.ab1), Chromas Lite software was used to visualize these raw sequences. The first and last 20 bases of these sequences were deleted, and the sequences trimmed to maintain highest level of accuracy; this included the ’N’s which shows that the nitrogenous base in that position was unidentified as well as additional unwanted bases to reduce the noise. A base quality of at least 25 shown at the software was retained in the sequence. Moreover, sequences were searched for stop codons, eventually exporting them in the FASTA file format (.fas) for future utilization.

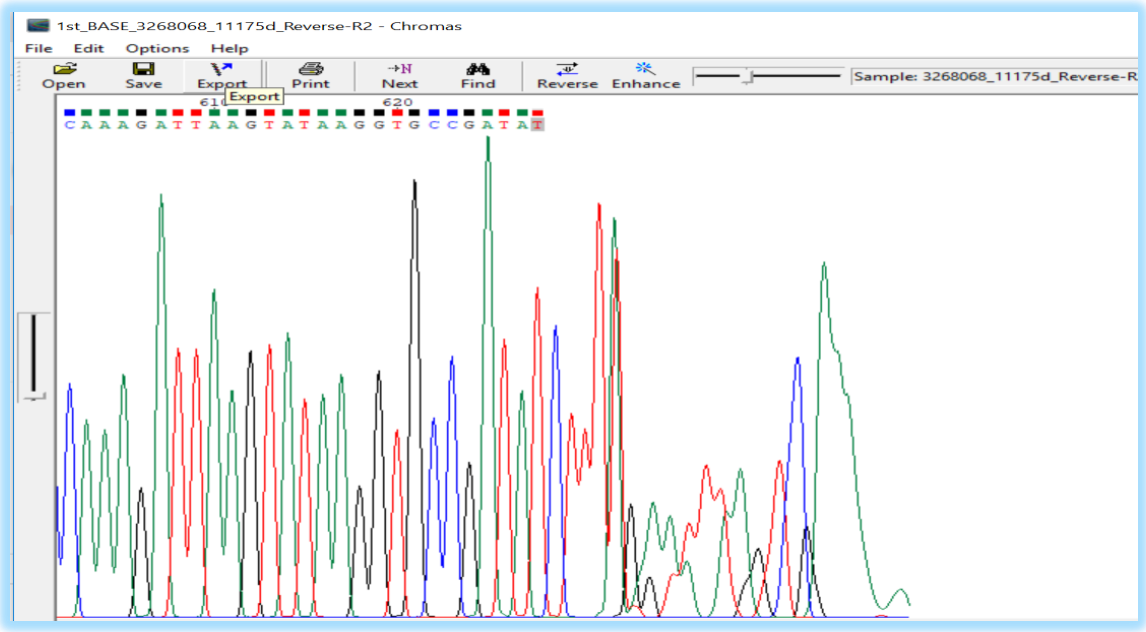

Figure 4. A portion of a COI gene sequence view in Chromas; around 20 bases from each end is deleted to maintain highest quality.

2.8 GenBank Submission

GenBank ® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. It is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis. The edited sequences, along with all the regarding information are submitted in GenBank which- upon request- withhold release of new submissions for a specified period of time. After submission, the sequences are crosschecked by professionals and confirmed if no further alteration is required. Within a period of 7-10 days, the Accession numbers of the respective nucleotide sequences are provided from NCBI, which are unique and subjected to just one sequence. If the sequences are not waiting to be published in print or online prior to the specified date, the sequence are released in the NCBI database for public access.

2.9 Accession numbers of sequences closely related to the study species

The CO1 nucleotide sequences of 18 species belonging to the Clupeiformes family were submitted in GenBank for obtaining accession numbers. The accession numbers were received in two weeks after successful submission of the sequences. Accession numbers of sequences closely related to the test organisms used in the analysis were noted and used for the next step.

2.9 Coalition of similar sequences from NCBI database

Against the 38 sequences belonging to three families of Clupeiformes, as many sequences as possible were queried against GenBank (Database resources of the National Center for Biotechnology Information. (2013), among the vast resources of Nucleotide database of the National Center for Biotechnology Information (NCBI). The Basic Local Alignment Search Tool (BLAST) algorithm was used to find their homology.

Though the NCBI database undergoes continual development, it was able to provide species matches of >97% sequence similarity for 90 of 91 samples tested. Not surprisingly, accurate species identification hinged on the known records within GenBank having correct taxonomic designations and being error-free. For purely organizational purposes of this study, a general rule that defined a top match with sequence similarity of at least 97% to indicate a potential species identity was used. This was done to compare the extracted sequences from Bangladesh, with sequences from the rest of the world, which would make the study more precise and accurate.

The top sequences producing similar and significant alignments were identified for the study species from NCBI and listed. In *Hilsa Kelee*, 16 similar hits were found in NCBI with a maximum identity of 100% and a minimum of 98%. As a general rule, a top match with a sequence similarity of at least 98% was required to be used as a criterion to designate potential species identifications.

Sequences were extracted from GenBank between June 2019 and August 2019 for all congeneric species pairs of animals possessing at least 400 bp of COI sequence from homologous sites. In total, 150 sequences met this criterion. Because their COI sequences were acquired using varied primers, they derive from different sections of the gene. In practice, most comparisons involved either a sequence block that extended from near the

5’ end of the gene to its middle or a block extending from the midst of the gene to its 3’ end.

2.10 MEGA (Molecular Evolutionary Genetic Analysis)

MEGA is an integrated tool for conducting automatic and manual sequence alignment, inferring phylogenetic trees, mining the web base data bases, estimating the rates of molecular evolution, and testing evolutionary hypothesis. Using this software, a total of 185 sequences (35 extracted, 150 taken from NCBI) were at first aligned using the Multiple Sequence Alignment program; in this case, MUSCLE (multiple sequence comparison by log-expectation). The type of program used and its accuracy depends on several factors related to the types of sequences being aligned: whether it is nucleic acid or protein, how related the families are, the to-be compared sequence lengths, the number of sequences, etc. MUSCLE, with its unique way of calculating distance measures (using kmer distance for an unaligned pair and Kimura distance for an aligned pair), progressive alignment using a new profile function called the log expectation score, and refinement using tree-dependent restricted partitioning, has the clear and distinct advantage of speed when compared to other programs like Clustal W, T-Coffee, ProbCons or COBALT.

2.11 Computation of Interspecies and Intraspecies distances

Later, inter and intra species distances were determined using the Kimura-2-Parameter, which required multiple sequence samples of each species. As a minimum of 3 to a maximum of 15 sequences were collected for each species beforehand using the BLAST hitlist, they were grouped according to **Families**, **Genus**, and **Species** in three separate rounds for comparing their Intra and Inter K2P distances.

The K2P distance calculates the number of differences accumulated between two sequences since their last common ancestor, as well as estimates the genetic distances among the sequences of alignment- for which MEGA builds a matrix with the pairwise distances. The overall notion was to find the intraspecies distance to be smaller than the interspecies distance, which remains one of the most important hallmarks of DNA barcoding process being accurately reliable.

2.12 %GC content analysis

The GC content was determined for each individual species, taking multiple sequences of the same species other than the sequences amplified to make sure the result was precise and fair. In this case as well, Mega X software was used for computation. Here, GC% for both the individual species and each of the three families were calculated for the 1st, 2nd and 3rd codon positions.

In the DNA barcoding journey, the GC% offers a subtle distinguishing test of the identity of unknown DNA, as well as the similarities between related sequences, due to the fact that the content of each of the three positions of the nucleotide sequences are compared, and undoubtedly, the GC-content throughout the genome differs distinctively between species.

2.13 Phylogenetic tree construction using MEGA

Phylogenetic relationships of genes or organisms are presented in a tree-like form known as phylogenetic tree. The branching pattern of a tree is called a topology. There are numerous methods for constructing phylogenetic trees from molecular data. While the neighborhood joining (NJ) method is preferred for accurate establishment of phylogenetic relationship and is an iterative clustering method based on the minimum-evolution criterion that produces an unrooted tree, UPGMA is an agglomerative hierarchical clustering method based on the average linkage method that produces a rooted phylogenetic tree. Since UPGMA method assumes equal rates of evolution, branch tips come out equal while as neighbor-joining tree method allows unequal rates of evolution, the branch lengths are proportional to the amount of change. However, the Maximum Likelihood method used in this case utilizes more complex evolution model, making the result closer to real life and takes into account the evolutionary model, showcasing ancestral history which is unlikely for UPGMA and NJ method as they establish relationships between sequences according to their genetic distance and constructs using phonetic method, not a phylogenetic one.

Here, both UPGMA and Neighbor Joining (NJ) method was generally implemented in case of barcoding.

Chapter 3

Results

3.1 Classification of Clupeiformes

35 new COI sequences from 18 different species found in Bangladeshi marine waters were generated for this study. With 150 sequences obtained from GenBank, a total of 185 samples were analyzed. These 35 sequences belonged to the three respective families Clupeidae, Engraulidae and Pritigasteridae. As it is crucial that fishes are generally studied by the Order they are members of, which is Clupeiformes alone in this case, the sequences were compared and studied in accordance to the families they belonged to. Out of the 18 fish species, 9 species (42%) belonged to **Engraulidae**, proclaiming itself as the largest representative Family of the study. On the other hand, 6 species belonged to **Clupeidae** while the rest 3 belonged to **Pristigasteridae**.

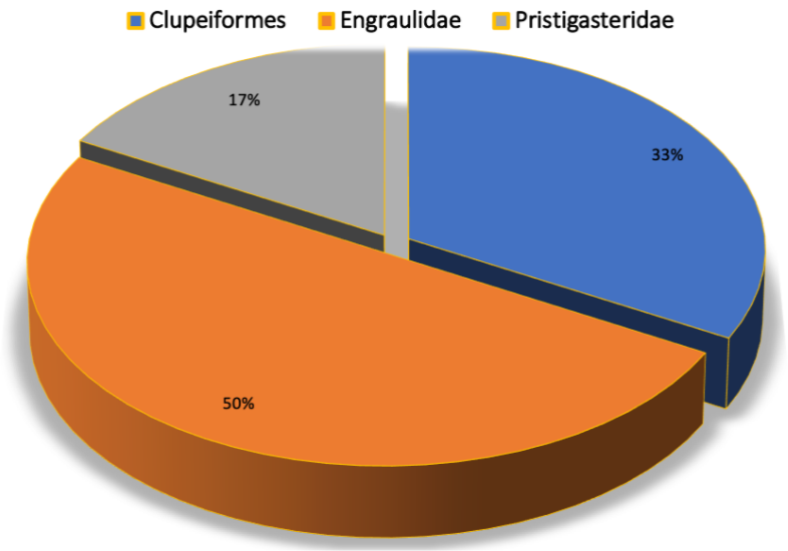

Figure 5. Pie chart showing percentage of each Genus among all extracted sequences used in the study

3.2 Observation of PCR - Amplified 655 bp of COI by Gel Electrophoresis

The 655 bp region of COI gene was amplified using PCR and the samples were separated in 1.2% agarose gel. Comparing the bands with the known molecular weight ladder, it was observed that PCR-amplified COI regions were situated between the 700bp and 500bp bands. This justified that the PCR amplicons were the desired 655 bp region of COI.

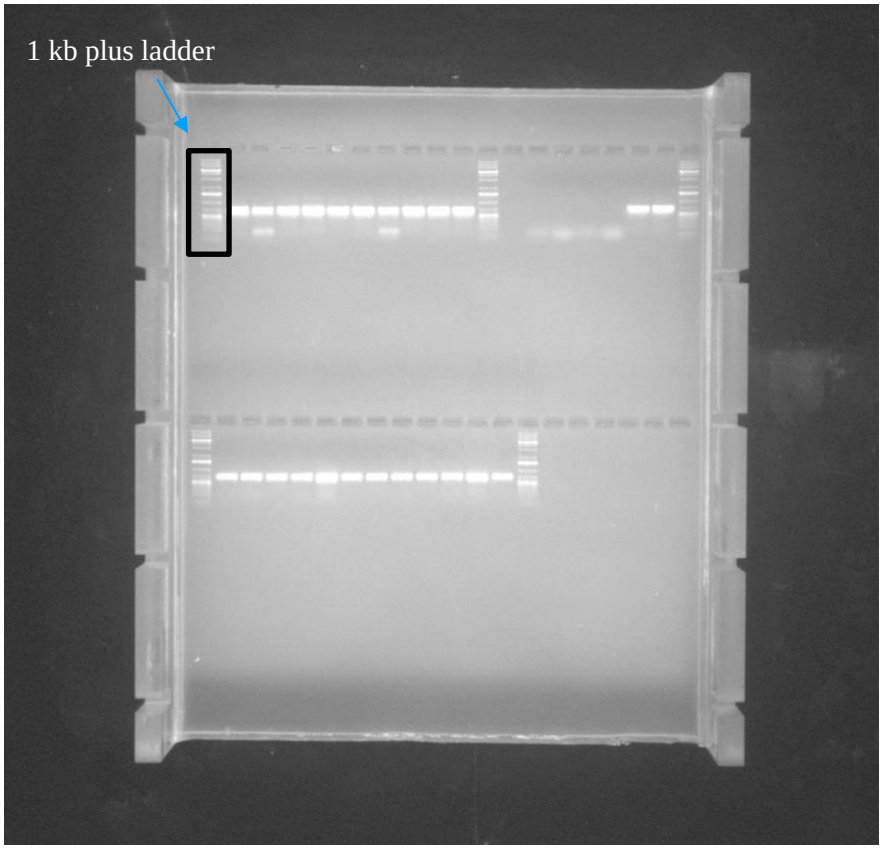

Figure 6 Electropherogram obtained after the electrophoresis of PCR products

3.3 G-C% of nucleotide sequences

		A-T%	G-C %	G-C% 1st position	G-C% 2nd position	G-C% 3rd position	Trend
1	*Anodontostoma_chacunda*	52.12	47.88	52.19	50.05	41.86	1>2>3
2	*Coilia dussumieri*	56.11	43.89	47.88	46.18	37.28	1>2>3
3	*Coilia ramacarti*	56.87	43.13	43.43	45.14	40.82	2>1>3
4	*Ilisha elongata*	55.48	44.52	46.49	46.32	40.53	1>2>3
5	*Opisthopterus tardoore*	57.17	42.83	46.04	44.71	37.52	1>2>3
6	*Pellona ditchela*	54.27	45.73	44.69	52.36	30.1	1>2>3
7	*Sardinella albella*	50.9	49.1	53.14	48.67	45.35	1>2>3
8	*Sardinella_ ongiceps*	51.09	48.91	52.98	49.99	43.59	1>2>3
9	*Setipinna melanochir*	54.91	45.09	48.84	45.38	40.86	1>2>3
10	*Setipinna phasa*	55.6	44.4	47.63	46.58	38.72	1>2>3
11	*Setipinna tenuifilis*	55.51	44.49	47.85	46.07	39.38	1>2>3
12	*Stolephorus dubiosus*	55.88	44.12	47.77	43.41	41.05	1>2>3
13	*Stolephorus indicus*	50.35	49.65	52.72	49.05	47.05	1>2>3
14	*Stolephorus waitei*	54.99	45.01	50.57	44.55	39.62	1>2>3
15	*Tenualosa ilisha*	51.5	48.5	52.8	47.88	44.58	1>2>3
16	*Tenualosa toli*	51.57	48.43	51.19	49.56	44.33	1>2>3
17	*Thryssa hamiltonii*	54.93	45.07	48.54	44.16	42.34	1>2>3
18	*Hilsa kelee*	48.31	51.69	54.74	52.94	47.31	1>3>2

Table 3. Showing the %GC content of the total sequence, as well as the first, second, and third codon position

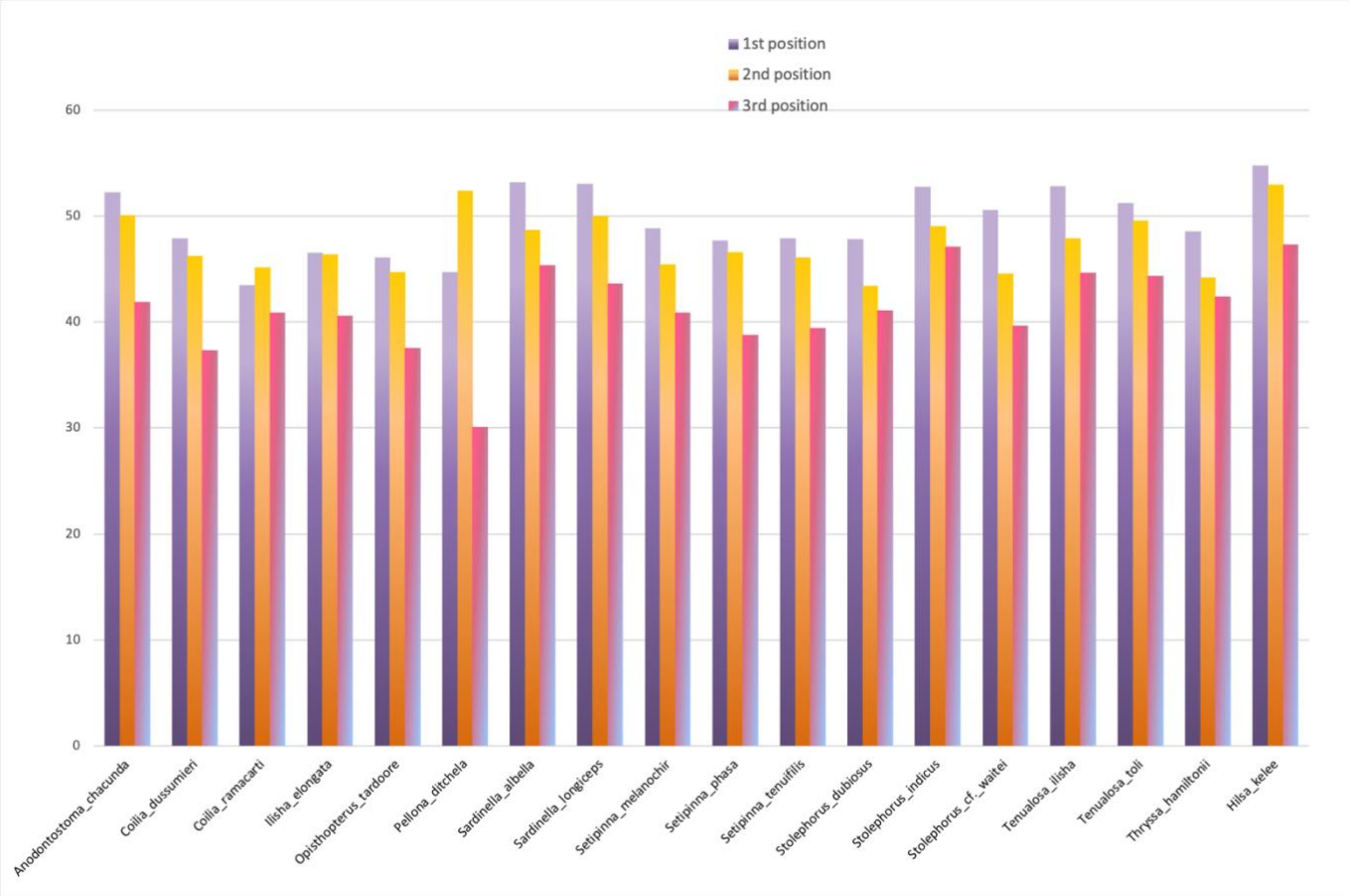

Figure 7 Showing the trend of %GC content in the first, second, and third codon position of different Clupeid species

The Pattern of %GC Content at Different Codons was $1^{st}>2^{nd}>3^{rd}$ for Each Species, with the exception of *Pellona ditchela* of the Pristigasteridae family and *Coilia ramacarati* of the Clupeidae family. This clearly proved that the trend of G-C content variation between the respective codons in case of Clupeiformes easily matched that of the sequence of vertebrate COI (44.7 ± 0.49). While the minimum overall value of %GC content was the observed for *Sardinella albela* (30.1%), that of *Setiphinna melanochir* was the highest (53.6%).

The GC content being a crucial parameter of a particular gene, it was used for comparing the genetic divergence between families as well.

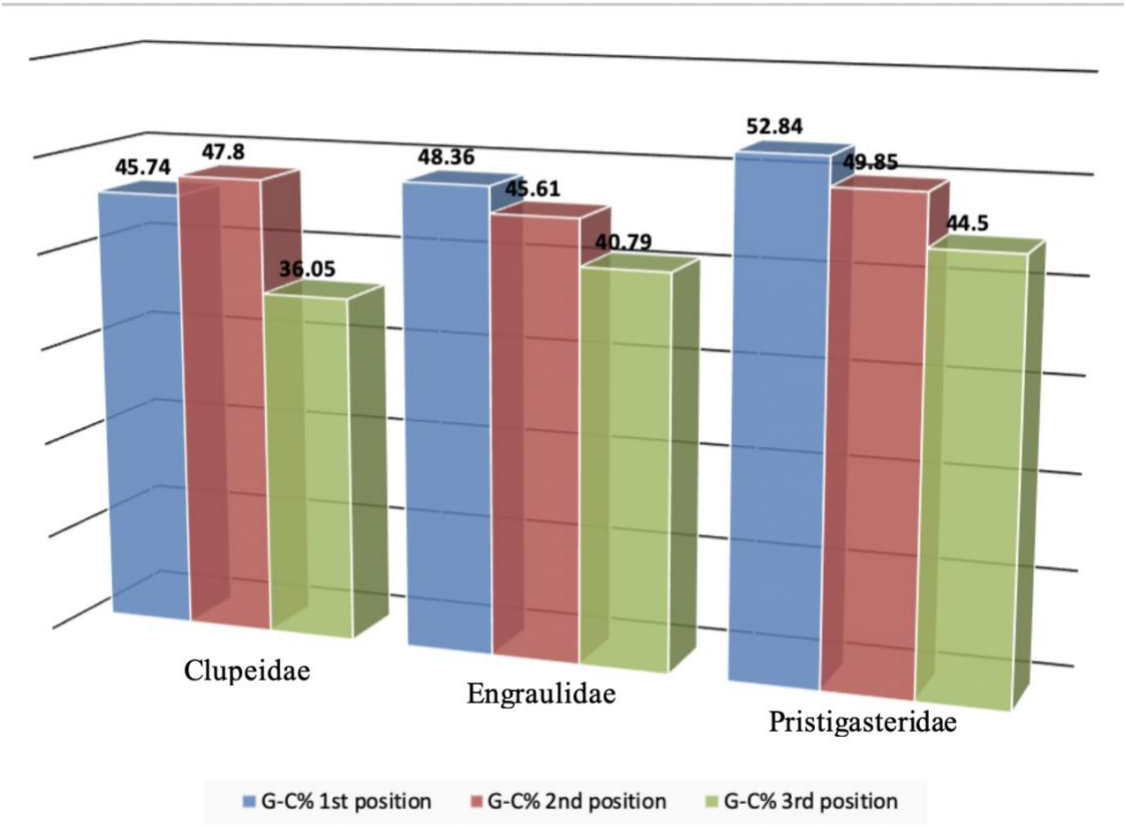

Figure 8. %GC content variation in 1st, 2nd and 3rd codon in case of Clupeidae, Engraulidae and Pristigasteridae

As mentioned before, *Coilia ramacarati* belonging to the Clupeidae family followed a different trend in comparison to the most of the sequences, which resulted to the overall GC% in the second position for the Clupeidae family to distinctively rise (47.8% at the 2nd position, which is greater than the 45.74% at the 1st). The other two families maintained an impressively prominent downtrend along the graph.

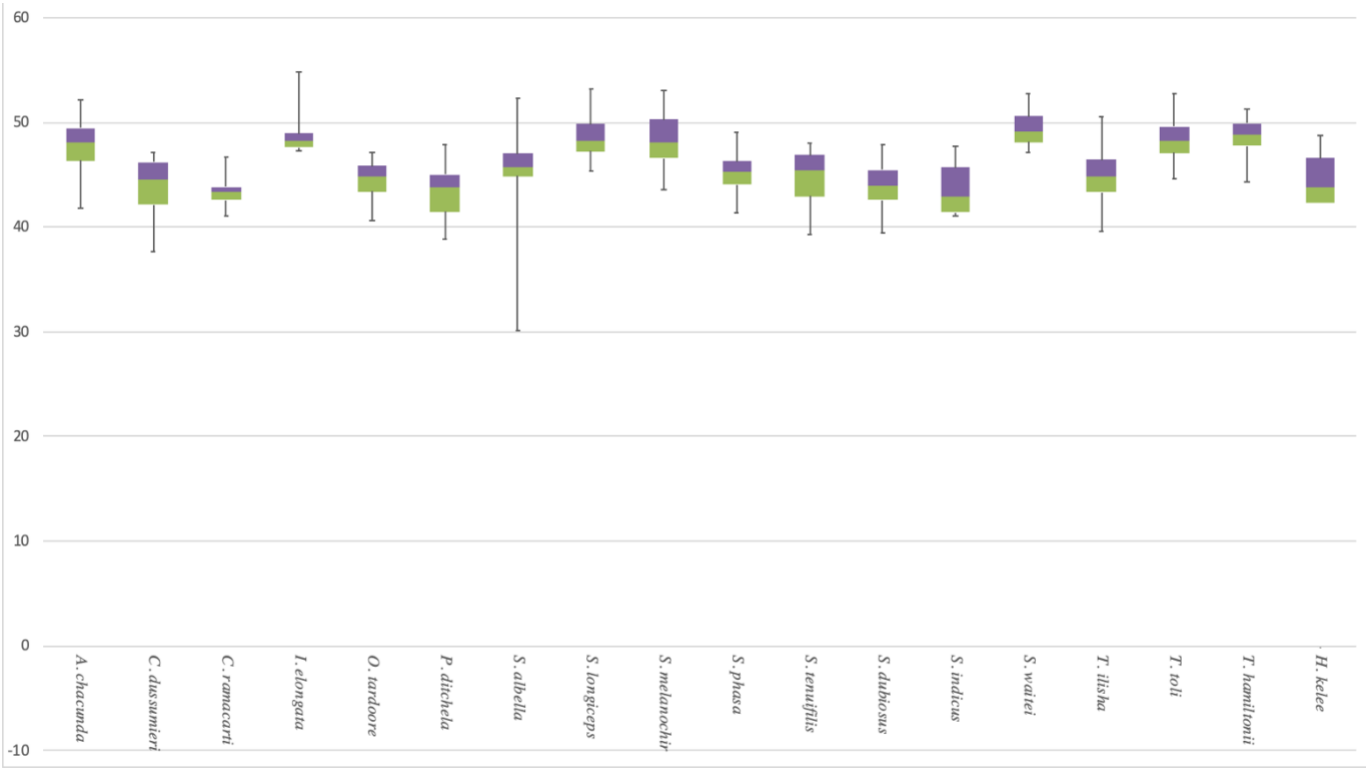

Figure 9. Box and whisker plot showing %GC content among species

Pellona ditchela) and 54.74 (*Hilsa Kelee*), with an average of 46.1%. While the interquartile range was mostly common in all species across the plotting area, one species, *Sardinella albela*, was found to have distinctively extending error bars, which depicted that the true value was quite far from the calculated value.

3.4 Intra and Interspecies distance

K2P Distance was Sharply Less in Intraspecies than InterSpecies

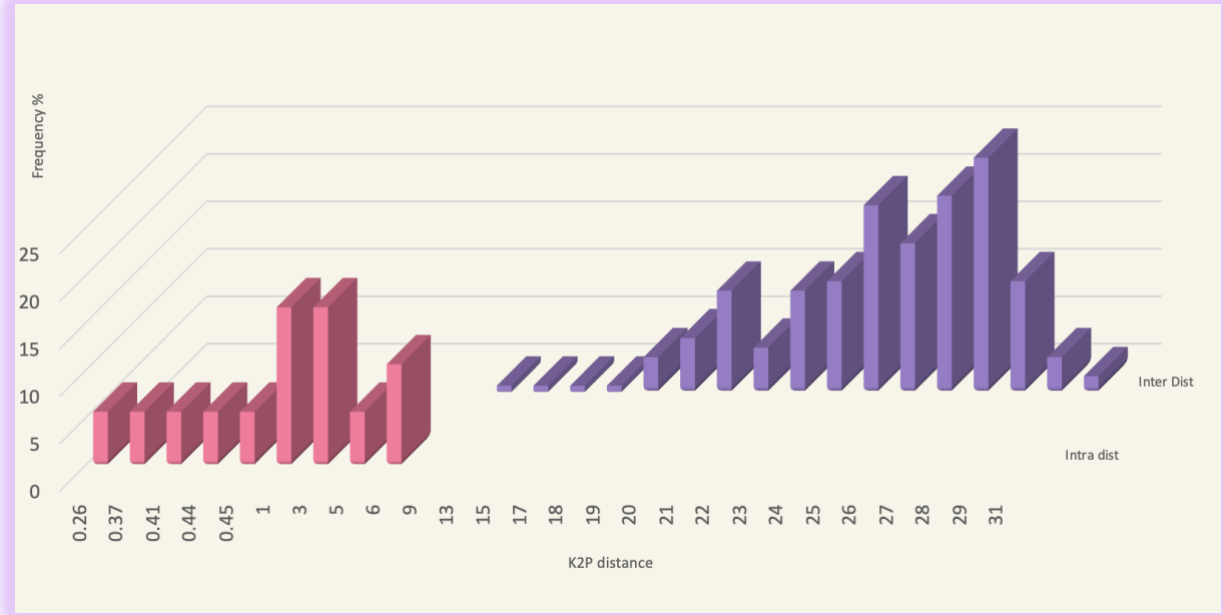

Figure 10. Frequency distribution of mean divergences for COI sequences for 185 samples, calculated by the Kimura 2- parameter model. Two taxonomic levels are represented: within species (pink bars) and between species (purple bars)

As expected, genetic divergence was observed to soar, with higher taxonomic rank—0.26% to 6.00% within species, whereas 9.00% to 29.56% between species. The difference between the intra and interspecies distance range was 3 units, which lies right at the middle of the expected range, which is 2-4 units. Since no category overlap was observed in any case, the distinguishing feature can be stated as quite distinctive (Figure 6).

However, recent studies have shown that closely related, but geographically distant species have been identified as the same species by DNA barcoding. In a study using mitochondrial COI and nuclear RAG1 markers, the method was incapable to distinguish between *Sardinella tawilis* and *Sardinella hualiensis*, two different species sharing the same genus. The interspecific K2P genetic distances based on the COI sequences ranged from 0% to 0.559% (mean = 0.286%), which fall below the 3–3.5% threshold to differentiate species (Hebert, et al., 2003[14]).

In addition, the low mean genetic divergence of <1% (0.191%) from RAG1 sequences between *S. tawilis* and *S. hualiensis* provided additional evidence in supporting the COI results, which suggested that *S. tawilis* and *S. hualiensis* belong only to a single species.

Table 4. Summary of genetic divergences of different taxonomic levels (based on the K2P distance model)

K2P Distance	*Taxa*	*Minimum*	*Mean*	*Median*	*Maximum*
Intra Species	18	0.26	1.88	0.45	6
Intra Genus	11	0	8.33	9	12
Intra Family	3	13	17	18	20

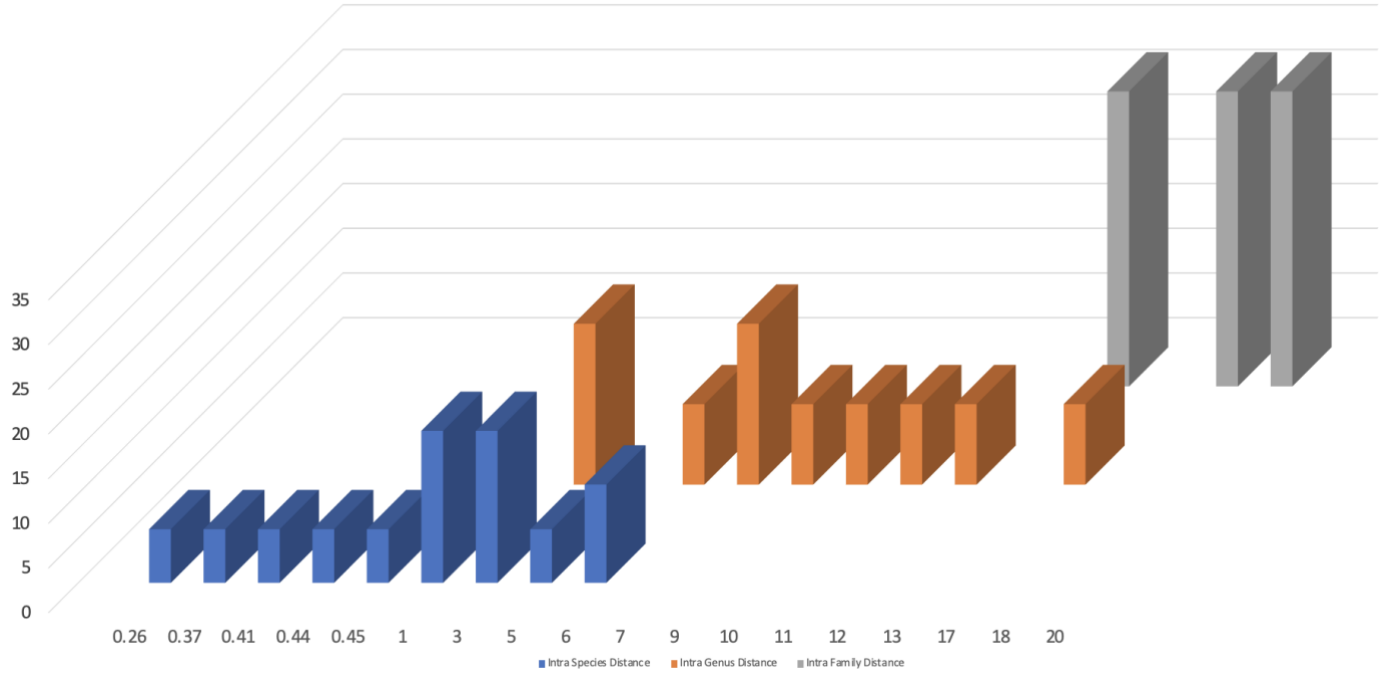

Figure 11. Frequency distribution of mean divergences for COI sequences for 185 samples, calculated by the Kimura 2- parameter model. Three taxonomic levels are represented: within species (blue bars), within genus (orange bars) and within family (grey bars).

It was observed that the level of divergence among congeneric species was about 4 times higher than among conspecifics, and divergence levels between confamiliars were about two times higher than congenerics. Thus, the mean conspecific, congeneric and confamiliar genetic distances were 1.88%, 8.33% and 17% respectively (Table 4).

The lowest intraspecific divergence was found among *Setipinna phasa* conspecifics (0.26%), while the highest belonged to *Opisthopterus tardoore* (6.0%), followed by *Stolephorus indicus* (5.0%). On the other hand, the highest interspecific genetic distance was between *Ilisha elongata* and *Stolephorus indicus* (31.47%) and the lowest between *Tenualosa ilisha* and *Tenualosa toli* (9.0%).

The congeneric divergence was 0% in case of Hilsa, soaring to 12.0% for Setipinna. In contrast, intergenus divergences ranged from 15.0% between *Ilisha* and *Opisthopterus*, and 28.0% between *Thryssa* and *Tenualosa*.

3.5 Phylogenetic analysis

The close phylogenetic relationships were found within the samples. This is clear evidenced in both Multiple Sequence Alignments and Molecular Phylogenetic analysis by Maximum Likelihood and NJ methods (Figure.10-13). The alignment was easy and clear because no gaps were found in the barcode sequences. In addition, no insertions, deletions or stop codons were observed in any sequence and lack of stop codons was consistent all along the result. The bootstrap value assigned to a node indicated how many times (%) the branches under a node stayed together. For example, if some two species had the node value 99, it meant out of 1000 bootstrap replicates, these two species remained under the same node together for 990 times.

Three phylogenetic trees were constructed to verify the distinction of three families of Clupeiformes and their evolutionary relationship. Three separate trees for the respective families are shown below (Figure.10-12), with a bigger tree showing all the families combined in one diagram (Figure.13).

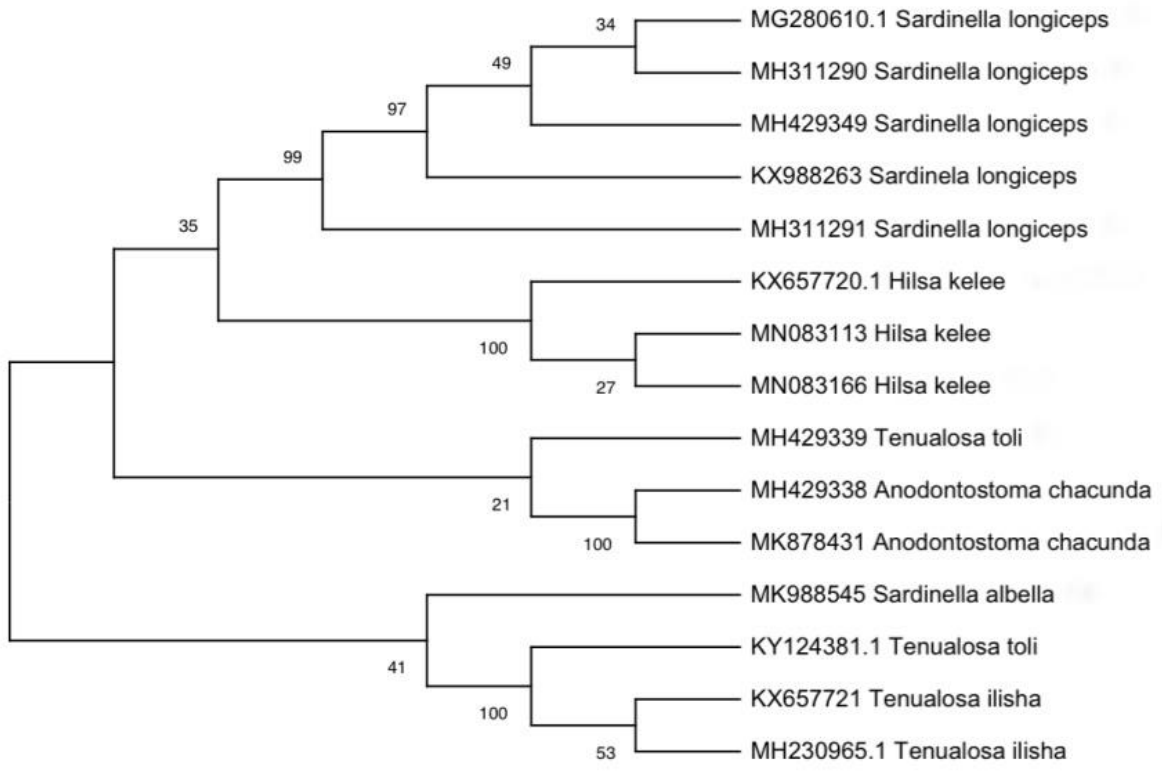

Figure 12. Phylogenetic tree showing species belonging to the Family- Clupeidae, using the Maximum Likelihood method with a bootstrap of 1000

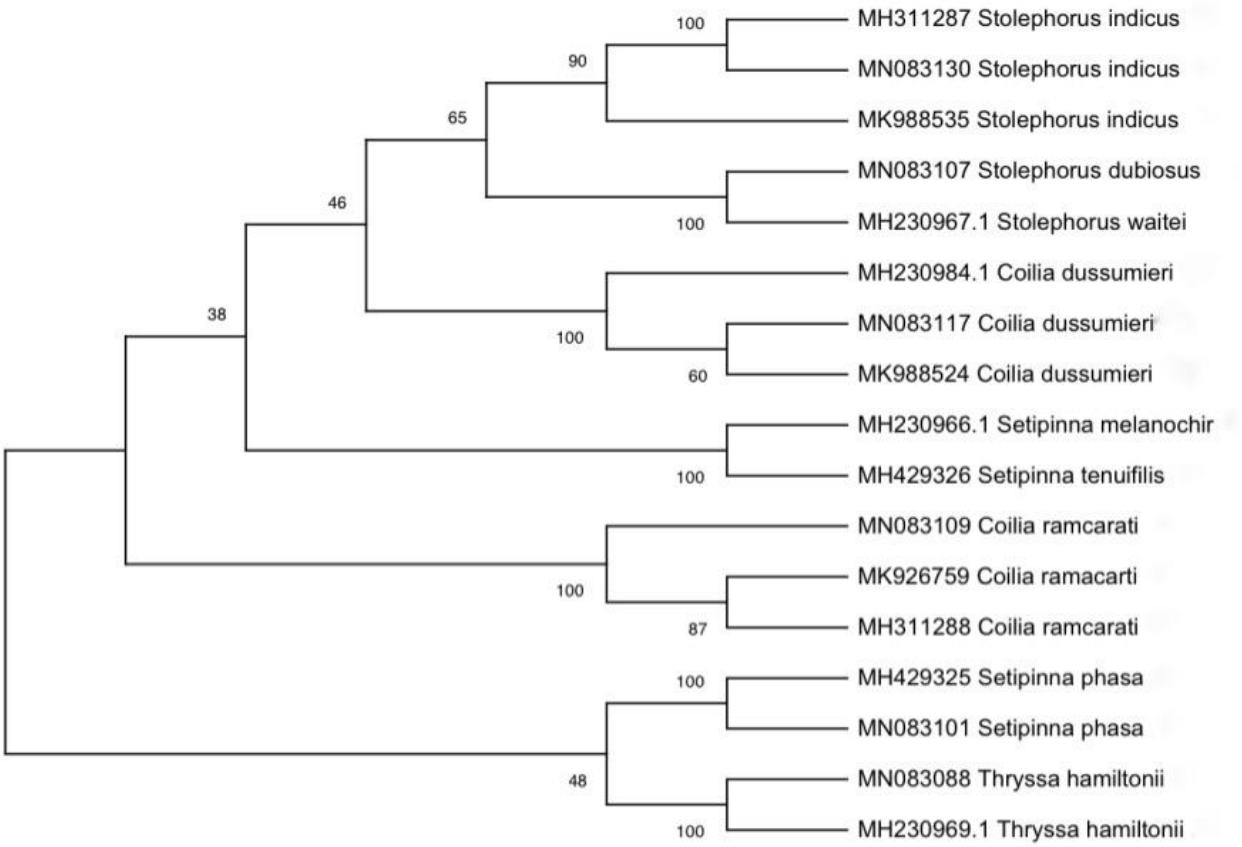

Figure 11. Phylogenetic tree showing species belonging to the Family- Engraulidae, using the Maximum Likelihood method with a bootstrap of 1000

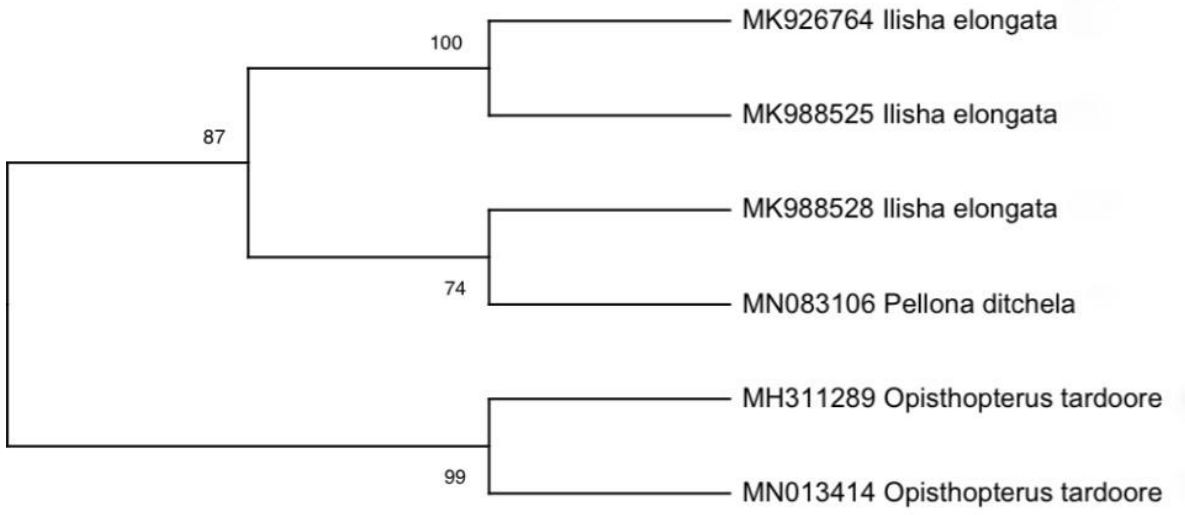

Figure 12. Phylogenetic tree showing species belonging to the Family- Pristigasteridae, using the Maximum Likelihood method with a bootstrap of 1000

In the above figures, all of the Bangladeshi Clupeidae fish species analyzed here exhibit a similar pattern of genetic diversity at COI, each being a single cluster of tightly related mtDNA sequences distinct from all other species. Therefore, the present survey fairly supports the view that the use of COI barcodes is a powerful tool for species identification.

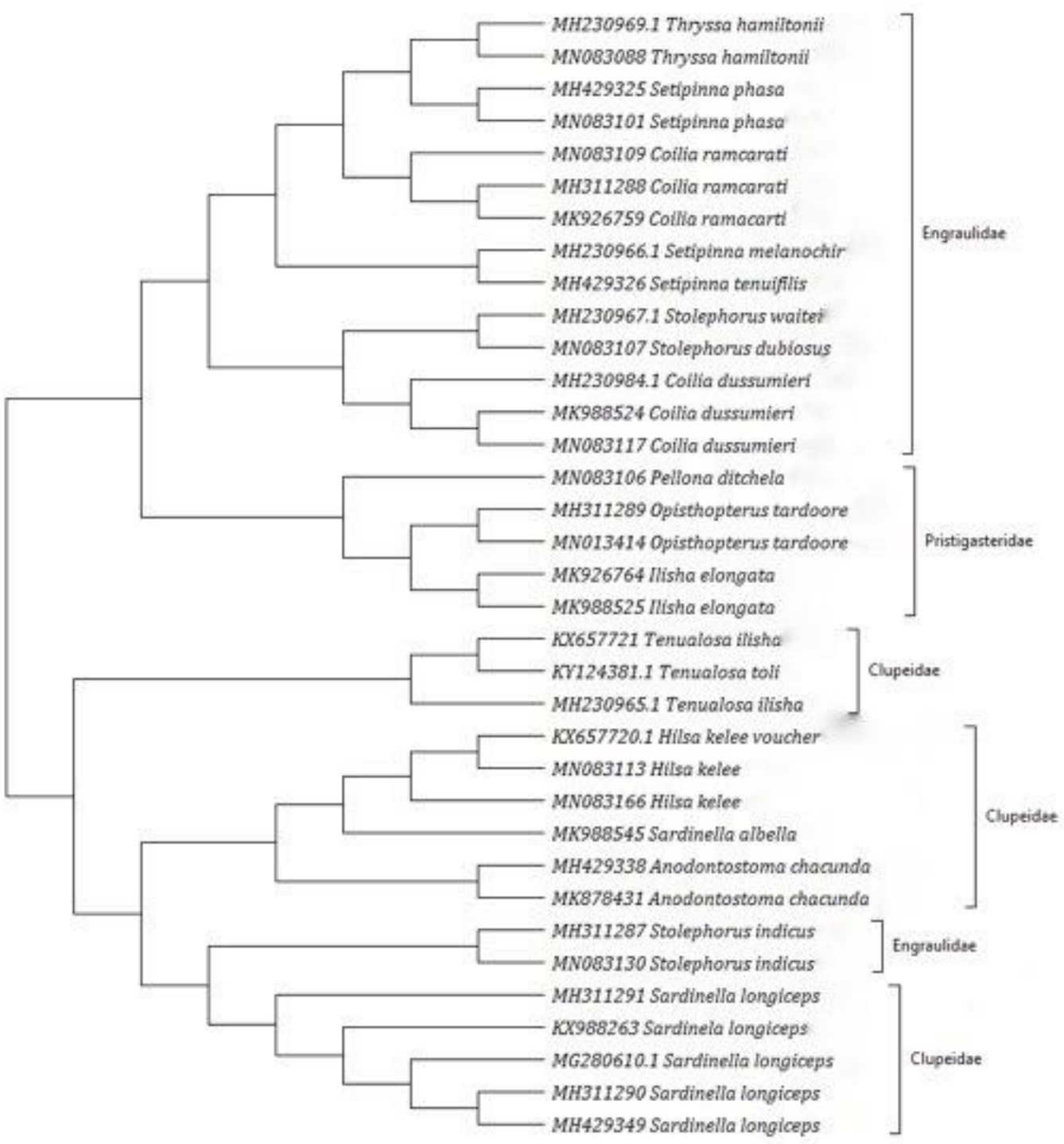

Figure 13. Molecular phylogenetic tree of Neighbour-Joining analysis (1000 replications) of three Clupeiformes families generated from partial COI mitochondrial sequences based on Kimura 2-parameter

In the figure, The NJ tree based on the K2P distance matrix essentially illustrates the relationships among the intrageneric units. The two genus Coilia and Setipinna were grouped together by their own common ancesters. The same happened for genus Stolephorus and Sardinella as well.

In all cases, species were clustered within their currently accepted families, distinctively separated in groups. While the Engraulidae family happen to have greater genetic similarity with Pritigasteridae than Clupidae, as shown in the tree, the Clupidae family forms its own cluster, distant to the rest two.

Chapter 4

Discussion

The three families belonging to the order of Clupeiformes (Clupeidae, Engraulidae, Pritigasteridae) being the main focus of this study, have their own reasons of importance. They share species of Ilisha, the most popular in Bangladesh, as well as other commercially important fishes, like chapila, shad, sardine, round sardinella, goldstrip sardinella, etc. The main objective of the whole study was to test and showcase the effectiveness of DNA barcoding in differentiating these species, as any form of misidentification or mislabeling can have detrimental effect on the commercial success of Bangladesh in the export industry.

Starting with the nucleotide content: the varying %GC content that strayed from the usual 1>2>3 position trend was not observed, except two species. Only two species, *Pellona ditchela* and *Coilia ramacarati*, showed a different pattern: the %GC content in the second position was the greatest of all, followed by the first and third position (2>1>3). While it was 52.36% in the 2nd position and 44.69% in the first in case of *P. ditchela*, the other species- *C. ramacarati*- had 45.14% GC content in the 2nd position, which was greater than the 43.43% in the 1st.

DNA barcoding is considered an appropriate tool for differentiating species in a group if interspecific variation exceeds intraspecific variation by one order of magnitude, known as the ‘barcoding gap’ (Martin & Fiedler, 2007[15]). The mean interspecific distance at 21.59% is more than an order of magnitude larger than the mean intraspecific divergence of 1.88%, demonstrating that the species examined in this study are suitable for DNA barcoding identification.

The level of genetic variation observed for the COI gene fragment was highly congruent with the taxonomic level. In general, the average conspecific, congeneric and confamilial K2P distances were within the range observed among the Canadian freshwater fishes (0.27%; 8.37%, and 15.38% respectively) (Hubert, et al., 2008[16]) and Australian marine fishes (0.39%, 9.93%, and 15.46% respectively) (Ward et al. 2005[17]) proving that the method is reliable. The character-based analysis results were completely in agreement with the genetic distance approach for species identification in all cases when both approaches were used. This supports the idea that both approaches used together can complement and facilitate species identification (Waugh, et al., 2007[18])

Overlapping of conspecific and congeneric levels of divergence was not seen. No genus or family showed interspecific divergences that fell below the threshold chosen to differentiate inter and intraspecific divergences. However, a similar argument cannot apply to the case of Rivulus, a recent research says. The two Cuban sister species *Rivulus cylindraceus* and *Rivulus insulaepinorum* showed a sequence divergence of 1.8%, similar to the 1.6% intraspecific comparisons of *R. cylindraceus*. Character-based analysis also failed to distinguish the two species and suggests a closer relationship between *R. insulaepinorum* and *R. cylindraceus* to be present in their habitat.

As evaluation of different standard parameters of COI sequences indicated that there were no significant changes in the COI sequences of the marine fish species investigated in our study, it is evident that all the values of the parameters taken into consideration were fairly consistent with the previous studies. In minor exceptions like the unexpected GC% trend in *Pellona ditchela* and *Coilia ramacarati*, we could not surely conclude only from studying the COI whether any genetic changes or reasons were responsible for this versatility. However, there is a chance that different chemicals and mutagens of the sea could be responsible to cause mutation both in COI as well as other genes in these two species. In this case, as these species are not available worldwide, it is believed that some unique genetic changes happen in these species by aquatic mutagens causing gradual decline could be easily put down.

Admittedly, this study is not sufficient in itself to solve those taxonomic issues. Inevitable limitations still persist, such as pseudogenes, an inability to detect hybrids, differences between gene and species trees, selection, geographic variation within species, the difficulty of differentiating recently derived species, etc (Rubinoff & Holland, 2005[19];

Hickerson, et al., 2006[20]; Frezal & Leblois, 2008[21]; Galtier, et al., 2009[22]). Such problems have already been recognized by DNA barcoding initiatives and are equally applicable to any mitochondrial marker. Nonetheless, as long as we endeavor to do more work in this area, DNA Barcoding has a prominent position in the field of Bioinformatics.

However, it serves to identify those taxa requiring further analysis, as well as providing the basic information on those species that may represent good models for comparative phylogeographic studies. A better knowledge of the systematics of Bangladeshi marine Clupeidae fishes may also contribute to improving the originality of the Hilsa business, one of the greatest commercial expertise in the country, as well as monitoring the impact of ecological changes. For example, the introduction and mislabeling fraud of morphologically similar fishes in the market has been a common practice in Bangladesh for years now. While customers spend a handful of money on buying what they think is genuine Hilsa, sellers and exporters are exploiting the fact that there is still a grey area in identifying the demanded species just by seeing how the fishes look like.

Chapter 5

Conclusion

The illegal mislabeling of fish and seafood species can have detrimental effects on both the industry and the consumer. To prevent these effects, which include economic fraud and health hazards, a research priority has been the development of species authentication techniques that are rapid, reliable, and reproducible. These include methods based on either species-specific/multiplex PCR or post-PCR analysis methods such as DNA sequencing, RFLP, SSCP, RAPD, and AFLP. Numerous nuclear and mitochondrial genetic markers have also been examined, with the most prominent being the mitochondrial gene cytochrome b. Methods that use smaller fragments, which can be analyzed in both raw and processed products, and the identification and quantification of species in mixed samples need to be optimized. In response to these challenges, future trends point to the use of technologies such as DNA microarray chips and quantitative real-time PCR methods. Furthermore, the use of databases has become increasingly important in this field by providing a compilation of genetic information on a variety of marine and freshwater fishes.

This paper provides a novel insight into how the sequence divergence varies between marine fishes, focusing on their genetic similarity in the trend when K2P is concerned with intra-and inter-species, genera and family distances, as well as percent GC content. It also shows a correlation between the nearest K2P distance and the percent GC content. While my work has begun to fill the gap in barcode coverage for Clupeiformes fishes in Bangladesh, there are many room for improvement in the whole scenario. While DNA barcoding has certainly gone from strength to strength and proved its accuracy quite impressively, there will always be considerable debate about how effective it really is.

Lastly, this study concentrates upon the commercially important Clupeid fish species and after having conducted this, it is believed that the issue of the preservation and professional identification of these fishes, particularly in front of the research community, has been highlighted. Also, it is believed that the articulation of information, a public awareness, careful investigation of mislabeling these species, inspiring the local fishermen about being aware of the mishandling of their product and whom they sell it to, needs to be taken special

care of. For implementing these, government of Bangladesh as well as NGOs have to play the most important role. It is sincerely trusted that this study will be able to inform the scientific community as well as the public and private organizations that it is high time we address this issue seriously.

Chapter 6

References

1. Rasmussen, R. S., & Morrissey, M. T. (2008). DNA-Based Methods for the Identification of Commercial Fish and Seafood Species. *Comprehensive Reviews in Food Science and Food Safety*, *7*(3), 280–295. doi: 10.1111/j.1541-4337.2008.00046.x

2. Woolfe, M., & Primrose, S. (2004). Food forensics: using DNA technology to combat misdescription and fraud. *Trends in Biotechnology*, *22*(5), 222–226. doi: 10.1016/j.tibtech.2004.03.010

3. Lenstra, J. (2003). DNA methods for identifying plant and animal species in food. *Food Authenticity and Traceability*. doi: 10.1201/9780203485385.ch2

4. Chapela, M., Sotelo, C., Calo-Mata, P., Perez-Martin, R., Rehbein, H., Hold, G., … Santos, A. (2002). Identification of Cephalopod Species (Ommastrephidae and Loliginidae) in Seafood Products by Forensically Informative Nucleotide Sequencing (FINS). *Journal of Food Science*, *67*(5), 1672–1676. doi: 10.1111/j.1365-2621.2002.tb08703.x

5. Bardsley, R. G., & Lockley, A. K. (n.d.). DNA based methods for meat authentication. *Special Publications Rapid Detection Assays for Food and Water*, 210–219. doi: 10.1039/9781847551818-00210

6. Bossier, P. (1999). Authentication of Seafood Products by DNA Patterns. *Journal of Food Science*, *64*(2), 189–193. doi: 10.1111/j.1365-2621.1999.tb15862.x

7. Civera, T. (2003). Species Identification and Safety of Fish Products. *Veterinary Research Communications*, *27*, 481–489. doi: 10.1023/b:verc.0000014205.87859.ab

8. Hebert, P. D., Ratnasingham, S., & Waard, J. R. D. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(suppl_1). doi: 10.1098/rsbl.2003.0025

9. Cespedes, A., Garcia, T., Carrera, E., Gonzalez, I., Fernandez, A., Asensio, L., … Martin, R. (2000). Genetic differentiation between sole (Solea solea) and Greenland halibut (Reinhardtius hippoglossoides) by PCR-RFLP analysis of a 12S rRNA gene fragment. *Journal of the Science of Food and Agriculture*, *80*(1), 29–32. doi: 10.1002/(sici)1097-0010(20000101)80:1<29::aid-jsfa470>3.0.co;2-4

10. Aranishi, F., Okimoto, T., & Ohkubo, M. (2005). A short-cut DNA extraction from cod caviar. *Journal of the Science of Food and Agriculture*, *86*(3), 425–428. doi: 10.1002/jsfa.2370

11. Carrera, E., Garcia, T., Cespedes, A., Gonzalez, I., Fernandez, A., Asensio, L. M., … Martin, R. (2000). Differentiation of smoked Salmo salar, Oncorhynchus mykiss and Brama raii using the nuclear marker 5S rDNA. *International Journal of Food Science and Technology*, *35*(4), 401–406. doi: 10.1046/j.1365-2621.2000.00404.x

12. Comesaña, A. S., Abella, P., & Sanjuan, A. (2003). Molecular identification of five commercial flatfish species by PCR-RFLP analysis of a 12S rRNA gene fragment. *Journal of the Science of Food and Agriculture*, *83*(8), 752–759. doi: 10.1002/jsfa.1368

13. Database resources of the National Center for Biotechnology Information. (2013). *Nucleic Acids Research*, *42*(D1). doi: 10.1093/nar/gkt1146

14. Hebert, P. D. N., Cywinska, A., Ball, S. L., & Dewaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), 313–321. doi: 10.1098/rspb.2002.2218

15. Martin, & Fiedler1, K. (2007, March 7). Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). Retrieved from https://frontiersinzoology.biomedcentral.com/articles/10.1186/1742-9994-4-8.

16. Hubert, N., Hanner, R., Holm, E., Mandrak, N. E., Taylor, E., Burridge, M., … Bernatchez, L. (n.d.). Identifying Canadian Freshwater Fishes through DNA Barcodes.

17. Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. N. (2005, October 29). DNA barcoding Australia's fish species. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1609232/.

18. Waugh, J. (2007). DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays*, *29*(2), 188–197. doi: 10.1002/bies.20529

19. Rubinoff, D., & Holland, B. S. (2005). Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. *Systematic Biology*, *54*(6), 952–961. doi: 10.1080/10635150500234674

20. Hickerson, M. J., Meyer, C. P., & Moritz, C. (2006). DNA Barcoding Will Often Fail to Discover New Animal Species over Broad Parameter Space. *Systematic Biology*, *55*(5), 729–739. doi: 10.1080/10635150600969898

21. Frézal, L., & Leblois, R. (2008). Four years of DNA barcoding: Current advances and prospects. *Infection, Genetics and Evolution*, *8*(5), 727–736. doi: 10.1016/j.meegid.2008.05.005

22. Galtier, N., Nabholz, B., Glémin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, *18*(22), 4541–4550. doi: 10.1111/j.1365-294x.2009.04380.x