DNA BARCODING OF SOME ENDEMIC CYPRINIDS OF ORNAMENTAL POTENTIAL

DILIP KUMAR., *B. F. Sc.*

Department of Fisheries Microbiology

College of Fisheries, Mangalore

Karnataka Veterinary, Animal and Fisheries Sciences University, Bidar

SEPTEMBER 2011

DNA BARCODING OF SOME ENDEMIC CYPRINIDS OF ORNAMENTAL POTENTIAL

Thesis submitted to the Karnataka Veterinary, Animal and Fisheries Sciences University, Bidar, in partial fulfillment of the requirements for
Degree of

MASTER OF FISHERIES SCIENCE

IN

FISHERIES MICROBIOLOGY

BY

DILIP KUMAR., ***B. F. Sc.***

Department of Fisheries Microbiology

College of Fisheries, Mangalore

Karnataka Veterinary, Animal and Fisheries Sciences University, Bidar

SEPTEMBER 2011

KARNATAKA VETERINARY, ANIMAL AND FISHERIES SCIENCES UNIVERSITY, BIDAR

DEPARTMENT OF FISHERY MICROBIOLOGY

COLLEGE OF FISHERIES, MANGALORE

****

CERTIFICATE

This is to certify that the thesis entitled **"DNA Barcoding Of Some Endemic Cyprinids Of Ornamental Potential"** submitted by **Mr. Dilip kumar.,** I.D No. **MFK-913** in partial fulfillment of the requirements for the award of Master of Fisheries Science in Fisheries Microbiology of the Karnataka Veterinary, Animal and Fisheries Sciences University, Bidar is a record of bonafide research work carried out by him during the period of his study in this University under my guidance and supervision and the thesis has not previously formed the basis for the award of any degree, diploma, association ship, fellowship or other similar titles.

(Dr. M.N. VENUGOPAL)

Professor
Department of Fisheries Microbiology

MANGALORE
SEPTEMBER 2011

APPROVED BY:

Chairman: _____
(Dr. M.N. VENUGOPAL)

Members: _____
(Dr. (Mrs.) I. KARUNASAGAR)

_____
(Shri. M. M. MARAGAL)

Dedicated to my
Lovely
Family

ACKNOWLEDGEMENTS

I wish to express my deep sense of gratitude and heartfelt thanks to my major advisor ***Dr. M. N. Venugopal****, Professor Department of Fisheries Microbiology and* ***Dr. (Mrs.) Indrani Karunasagar*** *Professor and Head, Department of Fisheries Microbiology who, with their pertinent comments and kind counsel, spent their precious time in advising and supporting me throughout the period of my research work. It has been a great pleasure to work under their guidance.*

I am extremely grateful to ***Dr. I. Karunasagar****, Senior Fishery Industry Officer (Quality) FAO, Rome, for his guidance and support rendered to me during the course of this study and preparation of the thesis.*

My wholehearted thanks and respect to ***Shri. M. M. Maragal,*** *Associate Professor, Department of Biochemistry, for his kind support during the course of this study.*

I sincerely thank ***the Dean, Dr. K. M. Shankar,*** *College of Fisheries, and also the University for providing me the necessary facilities in the institution to carry out my research work.*

I am deeply indebted to ***Dr. Ashwin Rai, Praveen Rai, Naveen Kumar Singayyah and Deekshit Vijay Kumar*** *for their immense help, constant and untiring support at every stage of my study. My heartfelt thanks to Dr. Anshu and Mr. Marcus Knight for their kind cooperation in this study.*

No words to express my gratitude to Nithin Suvarna, Rajeish Moger, Rajesh kumar, Santhosh, and Sampath for their friendship and timely support. I express my sincere thanks to Dr. Manja Naik, Dr. Malathi Shekar, Mahesh Shetty, Mohan S.A, Krishna Kumar, Safeena, Dr. Vijay, Devanand, Bishwajith, Patit, Debhashish, Mohammed, Dechamma, Akhila, Tanmoy, Mohan, MadhuKrishna, Mohini and all the non teaching staff members of the department and the college for the timely and valuable help provided to me.

I express my deep sense of gratitude towards my bosom friends, Girish, Lokesh, Harshit, Nagaraj Bhat, Olwyn, Raghavendra and all the friends from the university.

I express my sincere thanks to my sister, brother in law and my sweet little Sameeksha for their care and concern. I am grateful to my parents ***Mr. Y. B. Jogi*** *and* ***Mrs. Shantha. B. Jogi****, for their constant care, affection and support in all positive ways. Above all, I thank my brother, friend, role model and mentor,* ***Shrikantha Jogi*** *who is the real guiding force behind me, for being with me and supporting me always.*

Dilip Kumar

Mangalore
September, 2011

CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

The global biodiversity is being lost at an alarming rate and most part of the biodiversity remained unknown because the conventional morphological methods of cataloguing species are long, laborious and needs high level of expertise. However, cataloguing of all life forms becomes an important requirement in biodiversity studies where in taxonomy plays an important role. Conventional taxonomic classification of fishes is generally based on the morphometric characteristics, but often these characters fail to distinguish closely related species as well as eggs, larvae and young ones of a species. Hence, there is a need to use additional techniques which helps to distinguish the cryptic species. DNA barcoding is one such technique which helps to differentiate the fish species based on the DNA sequence.

Ornamental fish farming sector is a 16 billion dollar hobby based industry which is looking for new species of fish. According to FAO, India accounts for only 0.07% of the total world ornamental fish production. Though a large variety of marine and freshwater fishes with ornamental potential have been known from India, the ornamental fish industry is not fully exploiting these resources. Added to this, there are ambiguities that exist in the identification of some species of ornamental potential have limited the utilization of resource to its maximum potential in India. There have been several instances of same fish species being referred by different names and also different species called by the same name. Hence, there is an urgent need to bring more clarity to this confusion and give due credit to the respective species. The current study has been undertaken on four fishes of ornamental potential which are endemic Karnataka and Kerala which have identity conflict. In these regions, freshwater ornamental fish, *Puntius assimilis* and *Barilius canarensis* are often mistaken because of their close similarity with *P. filamentosus* and *B. bakeri*,

respectively. The present study is aimed at resolving this problem by using molecular approaches such as DNA barcoding.

Numerous fishes from the genus *Puntius* and *Barilius* have made significant contribution in the world ornamental fisheries. There is a widely debated problem in identifying these fishes as some of the species from each genus do not show much difference phenotypically. Conventional taxonomy which is based on morphological characters fails to distinguish between these closely related species. Hence, it is anticipated that DNA barcoding combined with the conventional taxonomy could help resolve the problem.

DNA barcode is an added tool to taxonomists who are trying to discover, distinguish and describe new species, and to anyone who is trying to assign an unidentified specimen to a known species. DNA barcodes can also be used by people who are not taxonomic experts on a particular group of organisms, and these barcodes can be used to identify specimens that are hard or impossible to identify with traditional methods (like damaged, incomplete, or immature specimens). It relies on the use of a standardized DNA region as a tag (marker) for rapid and accurate species identification. Nevertheless, DNA barcoding is not a new concept and the term 'DNA barcodes' was first used in 1993, in a paper that did not receive much attention from the scientific community. However, the golden age of DNA barcoding began in 2003.

After the initiation of Consortium for the Barcode of Life (CBOL) during the year 2004, which aims to promote global standards and coordinate research in DNA barcoding, it was decided that there should be a unique genetic barcode through which all the species on the planet earth could be identified. For this, a short DNA fragment such as Cytochrome c oxidase sub unit I (COI) which is a protein coding gene found in the mitochondrial genome and unique for all eukaryotic organisms was found suitable for the species identification

(Hebert *et al.*, 2003). According to Hebert *et al.,* (2003) barcoding includes selecting a short DNA sequence from a uniform locality in the genome of an organism. Besides taxonomists, the barcoding technique can be useful for scientists from other fields such as forensic science, biotechnology, food industries and animal diet.

Against this background, through an examination of the relevant type materials by fresh collections made in Kerala and Karnataka, an attempt is made to aid the taxonomy of these fishes by molecular technique, DNA barcoding, by harvesting and comparing the partial sequence of the gene Cytochrome c oxidase sub unit 1 (COI) as the barcode region in the present study with following objectives.

1. To use COI as DNA barcoding sequence for endemic cyprinids.
2. To differentiate endemic cyprinids based on barcoding data.

II. REVIEW OF LITERATURE

Cyprinid fishes are widely distributed in freshwater environment and include fishes that are important for food and ornamental purpose. Many cyprinid fishes of ornamental value are domesticated and bred under controlled conditions while, several species are collected from their natural habitat. Some fish species owing to their close similarity often misidentified and traded. The fish species with identity conflict include *Puntius assimilis*, *Puntius filamentosus*, *Barilius bakeri* and *Barilius canarensis*.

2.1. *Puntius*

The genus *Puntius* is currently includes over 100 species of small cyprinids. The need for full revision is being felt by many experts with the possibility of placing them into new or different genera. When describing the genus in 1822, Hamilton identified the defining characteristics as: "absence or presence of maxillary only or rostral and maxillary barbels; dorsal fin with last simple ray serrate or entire, branched rays usually 8; anal fin with last simple ray entire, branched rays usually 5; lateral line complete or incomplete, lateral-line scales 17-36 in row; cephalic cutaneous papillae minute or absent; pharyngeal teeth in 3 rows, usually 2,3,5/5,3,2; color pattern extremely variable. All the species currently in the genus are native to Southeast Asia, India and Sri Lanka. However, it was first identified and described as valid genera by Valenciennes in the year 1844.

2.1.1. *Puntius filamentosus* group

Puntius filamentosus group; an assemblage of similar-looking species that currently contains *P. arulius* , *P. exclamatio* (Present studies have shown that *P. exclamatio* could be a hybrid and not a true species), *Puntius filamentosus*, *P. singhala*, *P. srilankensis* and *P. tambraparniei* to

which *P. assimilis* was added recently by Pethiyagoda and Kottelat in their review of *Puntius filamentosus* group.

Expedition of the freshwaters of Kerala and in part, Tamil Nadu undertaken between 1991 and 1998 disclosed several probable species of filamented-fin barbs clearly distinguishable from *Puntius filamentosus* and one among them is *P. assimilis. P. assimilis* has been sold under the incorrect names *P. mahecola* or Mahecola barb. The problem regarding its true identity arouse in the year 2005 in the revision of the *P. filamentosus* 'group' by Maurice Kottlelat and Rohan Pethiyagoda.

Puntius assimilis was first identified by Jerdon in the year 1849 as a valid species and named as *Systomus assimilis.* But, in the subsequent revised classification, this species was considered as subspecies of *P. filamentosus* namely, *Puntius filamentosus assimilis.* Consequently it was considered as a geographical variant of *P. filamentosus.*

2.1.2 Taxonomic classification (Talwar and Jhingran, 1991)

Kingdom	: Animalia
Phylum	: Chordata
Sub phylum	: Vertebrata
Class	: Actinopterygii
Order	: Cypriniformes
Family	: Cyprinidae
Sub family	: Cyprininae
Genus	: *Puntius*

2.1.3 Habiat and distribution

Puntius assimilis

P. assimilis type fish is reported to be in various habitat-types in different rivers (Pethiyagoda and Kottelat, 2005). It is a schooling species and found in groups in its natural habitat. In the River Nethravati it was found to inhabit marginal areas with sluggish flow and muddy substrates. In the Chalakudy River in Kerala, it was collected from clear, rocky stretches that occur between the numerous waterfalls formed as the river travels down the Western Ghats. The fish in Kallada River were found in similar (clear water, rocky substrate) but slower-moving waters. These observations have clearly indicated that it is capable of living in different kind of habitats. According to Pethiyagoda and Kottelat (2005), *P. assimilis* type fish is endemic to the Southwest Indian states of Karnataka and Kerala where it has been collected from the Nethravati, Chalakudy and Kallada river basins in recent years. The type locality in the first description of the species was simply given as "a river in Canara" (Canara is a region of Karnataka comprising several districts).

Puntius filamentosus

This schooling species is most common in coastal floodplains. It is found in both fresh and brackish waters of rivers, estuaries, coastal marshes and reservoirs (Pethiyagoda and Kottelat, 2005). This fish species is endemic to the Southwest Indian states of Kerala, Tamil Nadu and Karnataka. The specimens are collected in their type locality close to the coastal town of Alappuzha (also known as Alleppey) in Kerala.

2.1.4. Taxonomic characteristics

Puntius assimilis

Adults (> 43.3 mm SL) of *P. assimilis* are distinguished from all other South Asian *Puntius* by a combination of the following characters: a black band about as wide as eye across each caudal-fin lobe (faint or absent in adults of the Kallada River population); lower lip continuous, a caudal blotch on 2-5 scales, commencing posterior to anal-fin origin; no prominent markings on body in advance of anal-fin origin. In adult males of the Chalakudy and Kallada River populations, branched dorsal rays prolonged into filament-like extensions. Additionally, adult *P. assimilis* may be distinguished from *P. filamentosus* by having inferior mouth (vs. subterminal) with maxillary barbels 23.5-33.3% of HL (vs. 2.8-8.1%).

Puntius filamentosus

Branched dorsal-fin rays prolonged into filament-like extensions (in adult males only); a black band about as wide as the eye near tip of each caudal-fin lobe; lower lip continuous; a caudal blotch on 2-5 scales, commencing posterior to anal-fin origin; no distinct markings on body in advance of anal-fin origin. Additionally, *P. filamentosus* is distinguished from *P. assimilis*, by having a subterminal mouth (vs. inferior); shorter maxillary barbels, 0.5-2.2% of SL (vs. 5.5-9.3%); post-orbital head length 11.0-12.1% of SL (vs. 8.7-10.4%); and interorbital width 11.2-12.2% of SL (vs. 10.0-11.1%).

2.1.5. The identity conflict

Francis Day, during the year 1875-1878, in his book, "*The Fishes of India*", considered *P. assimilis* as a separate species and named it as *Systomus assimilis*. But Jayaram (1999) in his book, "*The Freshwater Fishes of the Indian Region*", came up with a revised classification and included *S. assimilis /P. assimilis* in to *Puntius filamentosus* group considering that the debated

species is a subspecies of *P. filamentosus*. This has created difficulty whether to consider *P. assimilis* to be a separate species or a variant of *Puntius filamentosus*.

2.2. *Barilius*

These have a trout-like appearance and inhabit medium to fast torrential mountain streams, earning them the name 'hill-trouts'. On the rush of water they seek shelter under rocks and stones at the bottom or hide under crevices along the edges of the streams as they are affected by the swiftness of current though they are capable of progressing against it (Raagam et al., 2005). The body is moderately elongated and compressed sub cylindrically. Head is sharply pointed and abdomen is rounded. Snout is compressed, pointed, may be with pearl organs and tubercles. Mouth anterior and directed obliquely upwards. Upper jaw is a little larger than the lower. Pharyngeal teeth present in three rows. Dorsal fin with 7-12 branched rays and anal fin with 8-14 rays and caudal fin is forked. Eyes are large, superior in the anterior half of the head. Scales moderate and lateral line may be complete, incomplete or absent. Muscular pads are present in front of the bases of the pectoral fin. Air bladder is large extending to the entire length of the abdomen. Many species are beautifully colored with vertical bands or blotches or cluster of dots. They range in size from 4-12inches (Day, 1878).

2.2.1. Taxonomic classification (Talwar and Jhingran, 1991)

Kingdom : Animalia

Phylum : Chordata

Sub phylum : Vertebrata

Class : Actinopterygii

Order : Cypriniformes

Family : Cyprinidae

Sub family : Rasborinae

Genus : *Barillius*

2.2.2. Habitat and Distribution

***Barilius canarensis* (Jerdon's baril)**

Barilius canarensis inhabits clear streams with sandy and rocky bottom (freshwater) and are benthopelagic. In India, Jerdon's barils are found in Western Ghats and are endemic to Dakshina Kannada.

***Barilius bakeri* (Malabar baril)**

This species inhabits medium to fast torrential freshwater mountain streams. They seek shelter under the rocks and stones at the bottom or hide under crevices on the rush of water and hence benthopelagic. In India, it is reported from Western Ghats in Kerala and Karnataka and is endemic to Trivandrum, Kerala.

2.2.3. Taxonomic characteristics

Barilius canarensis

The fish is with two or more rows of oval blotches on the body. Mouth moderate; jaws short, maxilla extends to below anterior-third of orbit; barbels absent. Dorsal fin inserted in advance of anal fin, about midway between the snout-tip and base of caudal fin. The scales are moderate with few radii; predorsal scales 15. The tubercles are large and well-developed on head. Coloration: greenish above, golden on flanks, with a double row (rarely single row) of large vertical green oval blotches along the body. The fins are yellowish, with broad white margins.

Barilius bakeri

The fish is with vertical dots and grows up to about 15 cm in length. The body is silvery along the abdomen and grayish in the back; moderately elongated and compressed sub cylindrically. Head is sharply pointed and abdomen is rounded. Snout is compressed and pointed. Mouth anterior and directed obliquely upwards. Eyes are superior in the anterior half of the head and are large. Upper jaw is a little larger than the lower. Three rows of pharyngeal teeth present. The fins are greyish, with broad white margins. Dorsal fin with 7-12 branched rays and anal fin with 8-14 rays and forked caudal fin. Muscular pads present in front of the bases of pectoral fin.

2.2.4. The identity conflict

This fish species was first identified by Jerdon as a valid species during the year 1849. In his book, "*Fishes of India*", Francis Day described this to have 2 or more rows of oval blotches on the body. But in his book, "*Freshwater fishes of Indian Region*" Jayaram (1999) has misdescribed this species to be having 2 or more rows of large spots on the body which is now debated as a typographical error. But in India, Jayaram's book is widely used to identify the fishes and has resulted in the confusion and misidentification of this fish, the primary reason is due to the fact that variant of B. bakeri has been collected in Kerala and called as B. canarensis with two rows of dots.

2.3. Biodiversity and taxonomy

The biodiversity, defined as "the biological diversity among living organisms from all sources, including terrestrial, marine and other aquatic ecosystems, and the ecological complexes of which they are part" (International Convention on Biological Diversity, 1992), has emerged during the nineties as a topic of growing concern for sustainable

development. Taxonomy is the science that deals with the definition, diagnosis, description as well as naming of organisms and the subsequent organization of this information into systems of classification (Lipscomb, 2003). Species recognition is a basic need for large-scale biodiversity monitoring and conservation and the measuring of species prosperity is the most beneficial symbol of biodiversity. Initially, the majority of species were differentiated by their mature morphology but subsequently, more sophisticated approaches have been adopted. Electron microscopy, behavioral traits and biochemical markers have become the additional tools in the recent years for taxonomists for species description (http://www.barcoding.si.edu).

The cataloguing of species was founded more than 250 years ago by the Swedish naturalist Carl Linnaeus (1707-1778) who started the formal taxonomy by introducing the binomial nomenclature (including the genus and species name) which relied mainly on morphology, to describe the biodiversity (Linnaeus, 1756). This classical approach has not been able to record all species using morphological and behavioral observations and more recently biochemical markers, taxonomists were able to identify, describe and classify only a small fraction of the estimated species on earth. Though, around 1.7 million species have been described, the bulk of species on the earth remains still unknown and is predicted to vary widely, from 5 million to more than 100 million (Hawksworth and Kalin-Arroyo, 1995).

The global biodiversity is being lost at an alarming rate of 50-100 times the natural rate as a result of anthropogenic activities (Newmaster *et al.*, 2006). Further, most part of the biodiversity remained unknown because the conventional morphological methods of cataloguing species are long, laborious and needs high level of expertise (Hebert *et al.*, 2003a). In addition, the "morphological taxonomy" is found to be insufficient to record the

earth's biodiversity because of other three limitations. First, homoplasy (Vences *et al.*, 2005) and phenotypic plasticity to environmental factors (Saunders, 2005) of a given diagnostic character used for species recognition can lead to an incorrect identification. Second, this approach overlooks morphologically cryptic taxa, such as sibling species (i.e. morphologically same, but genetically different) that are common in many groups (Knowlton, 1993; van Velzen *et al.*, 2007). Third, as morphological keys are frequently effective only for a particular life stage or gender, it fails to recognize especially in their juvenile stages (Pegg *et al.*, 2006). Thus, the limitation of binomial Linnaean classification system has led to look for newer approaches of species identification methods for cataloguing the biodiversity. In this context, the DNA-based taxonomy is expected to provide a valuable informatory data to the conventional taxonomy and thereby meeting the rising need of truthful and handy taxonomic information (Tautz *et al.*, 2003).

2.3.1 DNA-based taxonomy

A distinguishing character of taxonomic significance is defined as "any attribute of a subject of a taxon that shows the difference with the subject of another taxon" (Ayala, 1983). It has long been known that DNA sequence diversity can be used to differentiate species since the nucleotide composition of the genome is specific for a given species (Manwell and Baker, 1963). The advantage of using DNA over proteins is because of its relatively stable nature which allows extraction from diverse types of samples, including museum specimens with damaged DNA, and from all life stages (Blaxter, 2004). In addition, DNA analyses are independent of the tissue origins (e.g. muscle, gonad, bone, etc.) because all types of cells contain same genetic information and the DNA information content

is higher compared to that of proteins, because of the degeneracy of the genetic code (Civera, 2003).

The use of a DNA- based system to examine evolutionary relationships was first applied by Carl Woese who documented the existence of the Archea domain by using the highly conserved 16S-rDNA gene coding for the small ribosomal subunit (Woese and Fox, 1977). Later, this approach was further used in numerous taxonomic groups with few morphological diagnostic characters as viruses, protists and bacteria (Nanney, 1982; Pace, 1997; Allander *et al.*, 2001). This technique, known as "DNA taxonomy", differs from DNA barcoding because it does not aim to tie up the genetic entities recognized through sequence analysis with Linnaean taxonomy and thus it is most functional for groups of organisms that lack meticulous taxonomic systems (Blaxter, 2004). In this regard, the development of an universal concept, "Molecular Operational Taxonomic Unit" (MOTU) (Floyd *et al.*, 2002; Blaxter *et al.*, 2005), for those organisms, such as meiofauna (Markmann and Tautz, 2005) or microorganisms was largely applied to illustrate clusters of genetic entities that are recognized entirely on the basis of the sequence similarity without any reference to the species name imposed with Linnaean binomial classification.

The DNA-based taxonomy system by detection of nucleotide sequence differences in a single gene for the identification of the organisms is thus an additional tool for assigning taxonomic status by matching the DNA sequence to a species already classified with Linnaean name but without giving it a central role (Godfray, 2002; Tautz *et al.*, 2003). This DNA- based system is not a replacement, but an add on tool for the conventional taxonomy. It has been recognized that the DNA sequence alone would not be sufficient to characterize a species (Ferguson *et al.*, 2002), except for some character-poor organisms such as soil

nematodes. However, an integration of broad range of phenotypic characters with molecular markers is expected to add strength to the species recognition (Dunn, 2003; Will *et al.*, 2005; Padial and De La Riva, 2009; Smith *et al.*, 2007). The introduction of DNA-based taxonomic approach involving the use of DNA as a new character for a taxonomic reference system was first proposed during the DNA Taxonomy Workshop held in 2002 in Munich, Germany.

2.4. The mitochondrial genome

The mitochondrial genome (mtDNA) (Fig.1) is a small circular genome and varies significantly among organisms in size, structure and gene content. It is a very handy molecular marker exhibiting significant characteristics and is most useful in evolutionary studies. In general, this genome is maternally inherited, with some exceptions of paternal or biparental inheritance (Korpelainen, 2004). Second, the entire genome represents a single linkage unit and that, along with its haploid nature, promotes the loss or the fixation of mtDNA haplotypes, reducing the diversity and thus sequence ambiguities from heterozygous genotypes within species, since it is non-recombining (Avise, 1989). Thus, the mtDNA allows the discrimination of even closely related species (Brown *et al.*, 1979; Juan *et al.*, 1996) as it generally evolves faster (about 5-10%) than single-copy nuclear genes at a rate of about 2% per million years in bilaterian metazoans (Ballard and Kreitman, 1995). This high evolutionary rate is due to regular mutations caused by elevated amount of reactive oxygen radicals produced through the respiratory chain that can chemically modify DNA, together with the absence of a dense protein-DNA complex that leaves mtDNA more accessible and, at the same time, more susceptible to damages caused by reactive oxygen radicals (Salgado *et al.*, 2008). The rate of evolution of the mtDNA is not consistent, but it displays variation in different regions. The maximum degree of conservation, with an average pairwise similarity over 75%, was found in

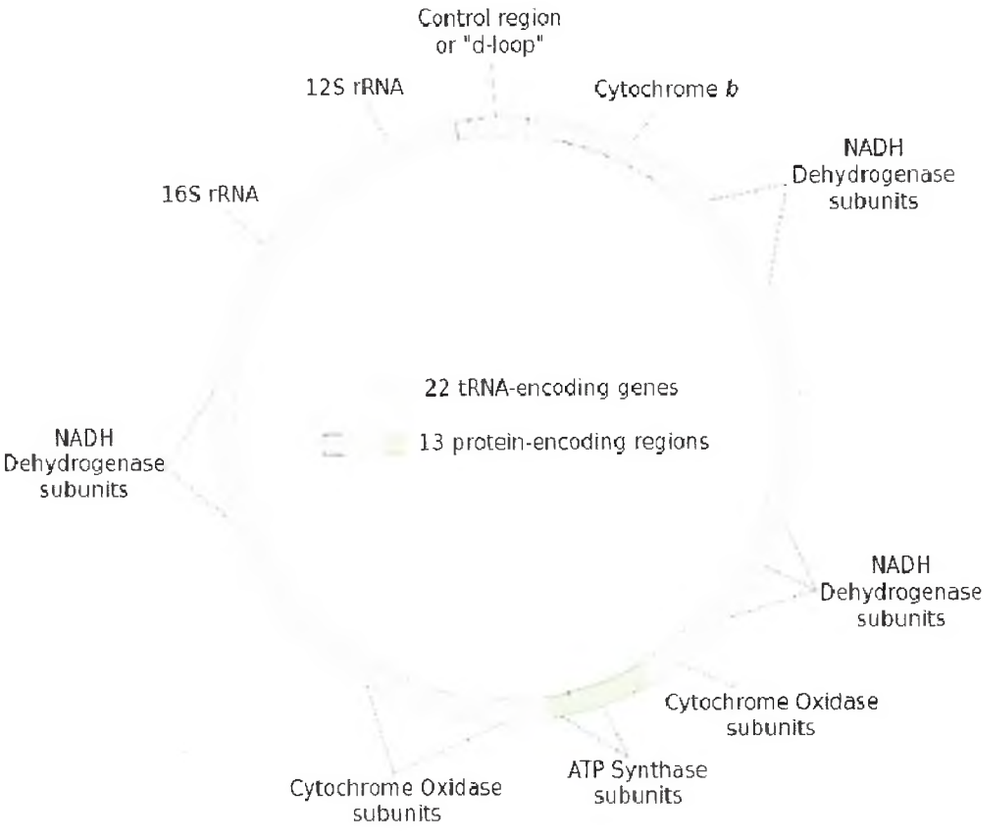

Fig.1. Mitochondrial genome showing different gene organization

the genes coding for the three subunits of the cytochrome c oxidase, the cytochrome b, the 16S rRNA and some tRNAs among all the functional regions in mammals (Saccone *et al.*, 1999). Generally the mutations are narrowed at third codon position, with high proportion of transitions than transversions, since it is less constrained by selection because of its four-fold degeneracy, since many mitochondrial genes are extremely conserved at the amino acid level (Hebert *et al.*, 2003a). Thus, the mutations generally are silent and selective neutral (Brown *et al.*, 1979), yielding many potentially phylogenetically informative characters. In addition, compared to the nuclear encoding single-copy genes, the mtDNA is present in multiple copies in the cell which improves the probability of amplifying template molecules even in case of highly degraded DNA, eg: processed food. Furthermore, sequence alignments of different species are much simplified as mtDNA lack introns and the low frequency of DNA deletions and insertions because sequence gaps are rare (Saccone *et al.*, 1999).

2.4.1. The ideal barcode marker

One of the foremost requirements of DNA barcoding is the selection of an ideal marker that easily discriminates any species in a given kingdom. Such markers have to meet certain criteria to be suitable for DNA barcoding of animals and plants. The best marker region should be orthologous between taxa, and amplifiable using universal primers, so as to standardize the procedure between wide ranges of taxa (Olmestead and Palmer, 1994; Taberlet *et al.*, 1996; Kress *et al.*, 2005). It should possess high conservation rate within species in order to generate the barcode gap (Hebert *et al.*, 2003a; Barrett and Hebert, 2005) and should also provide significant species-level genetic variability to identify species. It should be approximate of about 700 bp length so as to provide enough phylogenetically informative sites to easily assign species to its taxonomic group (family, genus, etc.), and also to allow PCR amplification and

DNA sequencing in one reaction (Shaw *et al.*, 2005). Further, the DNA barcode region should be simple to sequence without any long repeat regions; easy to analyze (i.e., more SNPs than insertions/deletions to avoid alignments ambiguity) and recoverable from degraded DNA samples, such as processed food, alcohol-preserved tissues stored in museums or forensic materials (Telechea *et al.*, 2005; Taberlet *et al.*, 2007).

The mitochondrial DNA (mtDNA) has been selected as the molecule for barcoding and use of the mtDNA to find out the evolutionary history within species was first suggested by John Avise (Avise *et al.*, 1987). Further, a large number of phylogenetic studies have been published using mtDNA and is widely considered as the first target genome in metazoans. However, the use of some mitochondrial genes encoding ribosomal DNA (12S, 16S) was not successful because of occurrence of frequent insertions and deletions that made the sequence alignment difficult (Doyle and Gaut, 2000). Then, the attention was focused on the protein-coding regions that provide the advantage of being arranged into codons. After comparing the 13 protein-coding genes according to their potential, COI (cytochrome c oxidase sub unit I) was proposed as suitable gene for DNA barcoding (Hebert *et al.*, 2003a). The sequence divergence at Cytochrome c Oxidase subunit I (COI) of mitochondrial genome is found to be highly discriminatory and enables the demarcation of closely allied species in all animal phyla except Cnideria (Hebert, 2003a). Using this approach it has become possible to identify eggs, larval stages and fragments of organisms which are not possible by conventional taxonomic method.

2.4.2. Cytochrome c oxidase subunit I

COI is 1,600 bp long, but only a partial sequence of 648 bp located near to the 5' end of the gene proved to be powerful in differentiating species and phylogeographic groups within

species. The COI gene was considered as the heart of the global bio identification system for animals as it shares all the criteria of an ideal barcoding marker (Chase *et al.*, 2005).

2.4.2.1. Advantages of using COI

The primary advantage of using COI is the recovery of the gene region from representatives of most animal phyla (Folmer *et al.,* 1994; Zhang and Hewitt, 1997) with no evidence of recovery of the nuclear pseudogenes (Hebert *et al.*, 2003a) using the universal primer pairs. Second, since the presence of insertions and deletions is almost absent and the evolution source is based on the nucleotide substitutions, the alignment of this region is easy (Hebert *et al.*, 2003a). Third, COI appears to have a higher range of phylogenetic signals than any other mitochondrial gene. Evolution of COI showed not only elevated rates of species discrimination (>95%) in vertebrate and invertebrate groups (Hebert *et* al., 2003b, 2004b), but also proved the ability to differentiate the phylogeographic groups within a single species (Lynch and Jarrell, 1993; Cox and Hebert, 2001; Wares and Cunningham, 2001). Further, in specimens whose DNA is degraded or where obtaining a full-length barcode is not feasible, the possibility of obtaining 'mini barcodes' (smaller fragments *i.e.* 100 bp of the standard COI barcode) have been found to be effective for species identifications (Hajibabaei *et al.*, 2006c).

2.5. DNA barcoding and conventional taxonomy

DNA barcoding is an additional tool that can be used along with conventional taxonomic studies for the precise identification of species and discovery of hidden species. But DNA barcoding may not guarantee an indisputable identification of recently diverged species that share alleles for a little time after the initial divide because of continuing gene flow. DNA barcoding is expected to refine species discovery once the barcode database is established as it enables a comprehensive phylogenetic study (van Velzen *et al.*, 2007). However, the

assessment of the species phylogeny through DNA barcoding is assumed to be conceptually incorrect as it is based on organellar marker that does not match to a gene for the speciation and hence it cannot monitor the evolutionary history of the taxa (Blaxter *et al.*, 2005). In spite of this drawback, the COI has been accepted as a best marker to assist the conventional taxonomic data in identifying a particular species.

2.6. DNA Barcoding of fish

Barcode of life is a recent development in the identification of species and a considerable progress has been achieved, since its inception by P.D.N Hebert, father of DNA barcoding. Ward *et al* (2007) have reported 99.5% species discrimination using COI barcoding data.

Identification of several fish species using barcoding approach has been reported by many investigators. Australian marine fish species were sequenced for a 655 bp region of mitochondrial COI gene by Ward *et al.,* (2005). The complete mitochondrial genome of the bullhead torrent catfish was sequenced by Kartavtsev et al., (2007). Mitochondrial DNA of the yellowfin seabream was completely sequenced and a genomic comparison was done among closely related species (Xia *et al.*, 2008). Identification of Canadian fresh water fishes through DNA barcode was reported by Hebert *et al.,* (2008). The sequence variation of COI of North American Coregonine species was used to develop a genetic assay based on COI (Schlei *et al.*, 2008). Analysis of the cytochrome c oxidase 1 gene sequence in six flatfish species (Pleuronectidae) of far- east Russia with inferences in phylogeny and taxonomy was reported by Kartavtsev *et al.*, (2008). Ward *et al.,* (2008) campaigned for DNA barcoding all fishes and need for establishing and strengthening Fish Barcoding of Life initiative (FISH-BOL). The ability of DNA barcoding to reliably identify the seven commercially important salmon and trout species was reported by Rasmussen *et al.*, (2009). Steinke *et al.,* (2009) worked on the

DNA based identifications for the import of ornamental fish into North America. The Northern blue-fin tuna was barcoded using COI, as per the requirement by Convention on International Trade Endangered Species (CITES) (Lowenstein *et al.,* 2009).

A full discrimination between all eight tuna species using the combination of two genetic markers, one mitochondrial and another nuclear was reported (Vinas *et al.*, 2009). The taxonomic boundaries and the limit of DNA barcoding in the Siberian sturgeon, *Acipenser baerii* was determined (Birstein *et al.,* 2009). In India also various institutions initiated the barcoding work among which National Bureau of Fish Genetic Resources, National Institute of Oceanography, Ministry of Science and Technology, Rajiv Gandhi Centre for Biotechnology, Center for Cellular and Molecular Biology, National Centre for Cell Science and Vector Control Research Centre (ICMR) were major participants. Species of marine fish covering Carangids, Clupeids, Scombrids, Groupers, Sciaenids, Silverbellies, Mullids, Polynemids and Silurids representing 79 Genera and 37 Families from the Indian Ocean have been barcoded for the first time using cytochrome c oxidase I gene (COI) of the mtDNA (Lakra *et al.,* 2009). Coral reef fish larvae were also identified through DNA barcoding (Hebert *et al.,* 2010). The DNA primers designed by Folmer *et al.*, (1994) for the amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates were successfully used for the DNA sequencing of a miniature cyprinid fish (Britz *et al.*, 2009). Thus, COI sequence has proved to be useful in the taxonomical identification of all organisms including the fishes without any bias.

2.7. DNA barcoding technical flowchart

The investigational steps of a DNA barcoding assay are very simple and include collection of sample, extraction of genomic DNA, amplification and sequencing of target region and creating the reference database.

2.7.1. Sampling and voucher specimens

The sequences have to be gathered from "holotype" specimens, *i.e.* original individuals stored in public collections (herbarium, museum, zoos, frozen tissue collections and other repositories of biological materials) or newly collected, which are identified by expert conventional taxonomists by means of morphological characters (Dalebout *et al.*, 2004). It is important to select new individuals with certain identities that should be stored as reference specimens. An identification voucher specimen and data such as images, locality information and ecological data associated with these specimens must be conserved as reference for future analyses. For smaller specimens that have to be completely sacrificed to extract DNA, morphological information can be conserved in the form of photograph (Tautz *et al.*, 2003). The need to preserve specimens guarantees the transparency of the database because it allows the reviews and re-analyses of a given sample at any time.

2.7.2. Extraction of genomic DNA

A tissue sample is taken from the collected specimens for DNA extraction by using suitable protocol. The extracted DNA is purified and the genomic DNA is stored in museum collections in desiccated or frozen condition for any subsequent amplification of additional genes (Blaxter, 2004).

2.7.3. Amplification and sequencing of specific target region

Once extracted, DNA serves as template from which the barcode COI, ITS, *matK* and *trnH-psbA* markers are amplified by PCR using universal primers (Folmer *et al.*, 1994). However, sometimes the development of taxon-specific primers and their combinations become necessary to obtain greater intra-generic accuracy (e.g. coral reef) (Neigel *et al.*, 2007). The obtained amplicons are then sequenced bidirectionaly and then manually checked and edited in order to validate sequence quality and detect eventual polymorphic sites, result of co-amplification of nuclear pseudogenes (Bensasson *et al.*, 2001).

2.7.4. Construction of reference database

Sequence information from the voucher samples are deposited in the database accessible from Barcoding of Life Database (BOLD) to allow unambiguous identification of specimens of unknown origin. The type specimen and the associated sequence becomes a reference record, only after the barcoding data are validated by the neighbour-joining method and by evaluating genetic distances within and between species.

2.7.5. Interrogation of barcode database

The identification step consists of submission of the 'query' sequence (COI sequence obtained from an unidentified sample), to the BOLD database through the BOLD-IDS in order to find the perfect match. BOLD-IDS accepts the DNA sequence from the barcode region and returns a taxonomic assignment to the species level, when possible, through the same sequence similarity search and the clustering method used for the validation step. In the case of COI marker, there are four different sequences subset in function of the validation of the sequences contained: only a subset of BOLD repository is a validated dataset because it includes sequence records with a sequence length of 500 bp, with a species level identification and referred to

many species represented by one or two individuals showing less than 2% sequence divergence. BOLD engine delivers a species identification providing the 20 closest matches, with a divergence value less than 1%, with the reference standard held within the database (Ratnasingham and Hebert, 2007). BOLD also generates a taxonomic identification summary and a NJ tree of species barcode sequences. Then the system can map specimen collection localities on a distribution map with high resolution and allows morphological comparison of voucher specimens when appropriate digital images are loaded. If the match is not obtained, the query sequence is assigned to a genus with a similarity divergence lower than 3%. Above all, if the unknown specimen does not match to any existing records in the barcode library, it should be flagged as a 'problem taxa' that deserves supplemental taxonomic analyses, rather than being discounted as a taxonomic error, suggesting that the sampling was not complete or may be presence of a new species, such as a cryptic species, or a new haplotype or geographical variant.

There are many advantages related to DNA barcoding. The result is reproducible at any time and by any researcher and therefore it represents a universally applicable method that can be linked to any kind of biological or biodiversity information. The experimental procedure of extracting DNA and amplifying specific markers is technically easy and usually does not require the destruction of the sample. The technique is fruitfull and effective in terms of cost and time and enables automated species identification, particularly useful in large sampling campaigns (Rusch *et al.*, 2007). The storage of DNA does not need particular attention because the molecule is very stable and any sample can be split into multiple subsamples, which can be sent to many museums as backups. Regarding the DNA sequencing step, technological progress is expected to guarantee a cheaper and faster way of sequencing (Tauz *et al.*, 2003).

III. MATERIAL AND METHODS

3.1. Samples

The *P. filamentosus* and *B. bakeri* samples (Fig.2-5) were collected from their type localities, back waters in Alappuzha and Kallar River in Trivandrum, respectively, in Kerala (Table.1). The *P. assimilis* and *B. canarensis* samples were collected near Ennehole from the upstream waters of river Swarna in Dakshina Kannada district in Karnataka. Following collection, the specimens were immediately placed in 95% alcohol and taken to the laboratory for the extraction of DNA. The species confirmation was done using conventional taxonomic techniques using suitable reference material out of which 5 fish each of four species were selected for barcoding.

3.2. Extraction of DNA

Approximately 40mg of muscle tissue was removed aseptically from the fishes. They were transferred to a sterile microfuge tube and homogenized. 950 µl of digestion buffer (containing EDTA, TritonX, gonidine hydrochloride and sodium acetate) was added and the homogenate was incubated at 4°C for 10 minutes.

Following incubation, the homogenate in 1.5 ml microfuge tube was centrifuged at 10000 rpm for 10 min. The supernatant of about 400-500 µl was transferred to a fresh tube following addition of equal volume of phenol chloroform iso amyl alcohol (PCIA), vortexed and centrifuged at 10000 rpm for 5 min. The aqueous phase about 300-400 µl was precipitated in a fresh tube by adding three volume of 100% ethanol and centrifuged at 14000 rpm for 10 min. The DNA pellet was washed with 1ml of 70% ethanol, centrifuged at 14000 rpm for 5 min and

Fig.2. ***Puntius filamentosus*** **collected from the backwaters of Alappuzha, Kerala.**

Fig.3. ***Puntius assimilis*** **collected from the river Swarna, Dakshina Kannada, Karnataka.**

Fig.4. ***Barilius canarensis*** **collected from the river Swarna, Dakshina Kannada, Karnataka.**

Fig.5. ***Barilius bakeri*** **collected from the river Kallar, Trivandrum, Kerala.**

then vacuum dried. Finally, pellet was dissolved in 100µl 1x TE buffer and stored at -20^{0}C for further analysis.

Digestion buffer:

10 mM Tris HCl	0.121 g
0.1 M EDTA	3.720 g
6 M Guanidium hydrochloride	57.30 g
0.1 M Sodium acetate	0.820 g
0.5% Triton X	0.5 ml

The above chemicals were dissolved in distilled water and the volume was made up to 100 ml.

PCIA:

PCIA was prepared by mixing Phenol, Chloroform and Iso amyl alcohol at a ratio of 25: 24: 1.

3.3. PCR assays

The PCR assays were performed to the DNA samples that were revived from 4^{0}C. Universal primers (Ward *et al.*, 2005) targeting *the gene Cytochrome c oxidase sub unit 1* were used. The PCR was performed in a 30 µl mixture consisting of 3 µl of 10X buffer (Bangalore Genei, Bangalore), 50 µM each of the four deoxynucleotide triphosphates (dNTPs), 25 pmol of each primer (Table 2), and 1.5 U of *Taq* DNA polymerase (Bangalore Genei, Bangalore, India). Two microliters of extracted DNA was used as template. The PCR assays were performed in a programmable thermocycler (MJ Research., USA). In all the reactions, an initial denaturation of template DNA at 95^{0}C for 2 min and a final elongation at 72^{0}C for 10 min were followed. The primers used and respective cycling conditions are mentioned in Table 3. The PCR

Table 1. Collection of Fish species from type locality.

Type Locality	Type species
Kallar river, Trivandrum	*Barilius bakeri*
Backwaters, Allappuzha	*Puntius filamentosus*
Swarna river, Karkala	*Barilius canarensis*
Swarna river, Karkala	*Puntius assimilis*

Table 2. Primers used for amplification of COI gene of *P. filamentosus, P. assimilis, B. bakeri and B. canarensis*

Primer	Sequence	Amplicon size (bp)	Reference
FISH F1 FISH R1	5'TCAACCAACCACAAAGACATTGGCAC3', 5'TAGACTTCTGGGTGGCCAAAGAATCA3',	655 bp	Ward *et al.* (2004)

Table 3. Thermocycling conditions for the amplification of *COI* gene of *P. filamentosus, P. assimilis, B. bakeri and B. canarensis*

Primer	Cycling condition			Number of cycles
	Denaturation	Primer annealing	Primer extension	
FISH F1,R1	94°C for 30 sec	54°C for 30 sec	72°C for 1 min	35

products thus obtained were resolved on 1.5 % agarose gel, stained with ethidium bromide (0.5 µg/ml) and visualized using a transilluminator (Herolab, Germany).

3.4. Detection of PCR products by agarose gel electrophoresis

Twelve micro liters of the PCR product was mixed with 4 µl of sample loading buffer and loaded on to agarose gel. The products were resolved on 1.5% agarose gels, at constant 130 V in 1 X TAE buffer, stained with ethidium bromide (0.5 µg/ml), photographed and analysed using a gel documentation system (Herolab, Germany).

TAE Buffer (50X)

Tris base	242 g
Glacial acetic acid	57.1 g
0.5 M EDTA (pH 8.0)	100 ml

The solution was made up to a final volume of 1 liter using millipore water.

Sample Loading Buffer (6X)

Bromophenol blue	0.25 g
Sucrose	40.0 g
Distilled water	100 ml

3.5. Purification of PCR products

To increase the volume, the gene was PCR amplified in bulk using the primers and cycling conditions mentioned earlier. The PCR products were purified before cloning to remove contaminants like *Taq* DNA polymerase, primer dimers, remaining dNTPs which may interfere with subsequent processes, using QIAquick PCR purification kit (Qiagen) (Fig.6).

To one volume of the PCR product, 5 volume of buffer PB was added, mixed and then transferred to QIAquick spin column placed in a 2 ml collection tube. The flow through was discarded after centrifugation at 10,000 x *g* for 1 min, washed by adding 0.75 ml buffer PE containing ethanol followed by centrifugation at 10,000 x *g* for 1 min. The DNA was eluted by adding 50 µl of elution buffer (10 mM Tris-Cl, pH 8.5) to the QIAquick column placed in a fresh microcentrifuge tube and collected by centrifuging the column at 10,000 x *g* for 1 min.

3.6. Cloning

3.6.1. Ligation of PCR products into the expression vector

The cloning vector pDrive (Qiagen, Germany) was used in this study for cloning of COI gene. The vector map is shown in the Figure 7. The ligation reaction was carried out according to the manufacturer's instructions. The ligation reaction mixture was prepared by mixing 4 µl purified PCR products (purified PCR product of gene COI) and 1 µl vector (50 ng/µl) in a 0.2 ml PCR tube. Five µl of 2x ligation master mix was added to each tube, mixed gently and incubated at 16°C in a thermal cycling block for 2 hours.

3.6.2. Transformation

Transformation was carried out by following manufacturer's instruction (Qiagen (2003). Frozen (80°C) competent *E.coli* strain Ez was used in transformation of COI. An aliquot of the ligation mix was transferred to tube containing 100 µl competent cells. After adding the ligation mix to the competent cells, the contents were mixed gently and incubated in ice for 30 min. The cells were subjected to heat shock at 42°C for 90 seconds in a water bath. The tubes were then immediately transferred to ice and cells were allowed to chill for 1-2 min in ice. Then 500 µl of super optimal culture (SOC) medium (maintained in room temperature) was

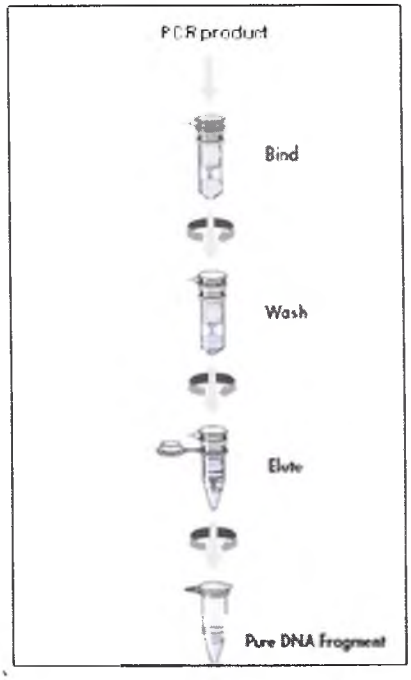

Fig. 6. Schematization of purification of PCR products using QIAquick spin column.

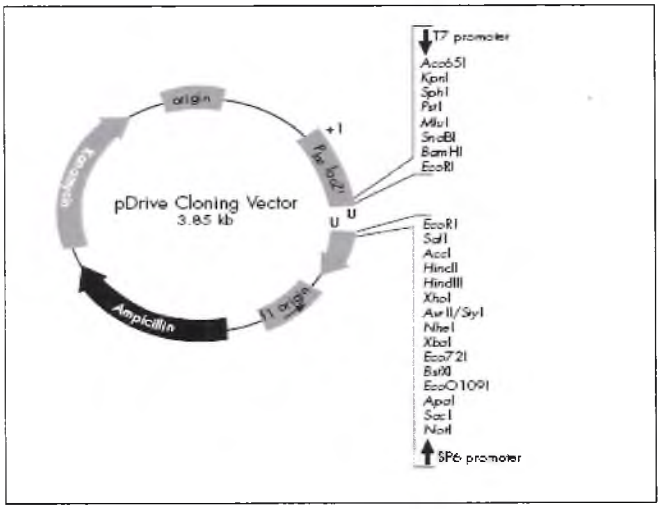

Fig.7. Schematization of PCR cloning in Qiagen PCR Cloning pDrive. Reproduced from Qiagen PCR Cloning Kit Instruction manual.

added to the tubes and incubated at 37°C for 2 h with vigorous shaking (200 rpm) in horizontal position. 50 µl and 100 µl of the transformed broth were spread on prewarmed LB agar plates containing antibiotics (100 µg/ml ampicillin and 25 µg/ml kanamycin). The plates were incubated overnight at 37°C.

LB agar with antibiotics

Four grams of dehydrated LB agar was dissolved in 100 ml of distilled water and sterilized at 121°C for 15 min. When the medium was cooled to 50°C, antibiotics were added from the stock solutions to obtain the desired concentrations.

Ampicillin stock solution (100 mg/ml)

Ampicillin stock solution was prepared by dissolving 0.1089 g of ampicillin powder (Calbiochem, USA) having 98% purity in 1 ml distilled water. The solution was filter sterilized using 0.22 µm syringe filter (Pall Corporation, USA) and stored in aliquots at -20°C.

Kanamycin stock solution (25 mg/ml)

Kanamycin stock solution was prepared by dissolving 0.0385 g of kanamycin powder (HiMedia, Mumbai) having the assay potency of 650 µg/mg in 1 ml distilled water. The solution was filter sterilized using 0.22 µm syringe filter (Pall Corporation, USA) and stored in aliquots at -20°C.

SOC medium

Tryptone	2%
Yeast extract	0.5%
NaCl	10 mM
KCl	2.5 mM
$MgCl_2$	10 mM

$MgSO_4$	10 mM
Glucose	20 mM

All the components except glucose were dissolved in 100 ml distilled water, sterilized by autoclaving at 121°C for 15 min. When the medium is cooled, filter sterilized glucose was added and stored at 4°C.

3.6.3. Preparation of competent cells

Competent cells were prepared according to the supplier's protocol (Qiagen, 2003). A loopful of *E. coli* strain was removed from the supplied vial and streaked on LB agar plates containing antibiotics (100 µg/ml ampicillin and 25 µg/ml kanamycin) and incubated overnight at 37°C. A single colony was picked and inoculated into 10 ml of LB broth containing antibiotic and incubated overnight at 37°C. One ml of the overnight grown culture was added into 100 ml of pre warmed LB broth containing antibiotic in a 250 ml flask and incubated for 90-120 min with shaking at 200 rpm at 37°C. When OD_{600} reached 0.5, culture was cooled on ice, transferred to 50 ml polypropylene tubes and centrifuged at 4000 × *g* for 10 min at 4°C (Heraeus, Germany). The supernatant was discarded and the pellet was resuspended in 30 ml of ice cold TFB1 buffer (30 ml for a 100 ml culture), incubated on ice for an additional 90 min and centrifuged at 4000 × *g* at 4°C for 10 min. The supernatant was discarded and cell pellet was resuspended in 4 ml of ice cold TFB2 buffer. 100 µl of suspension of competent cells was aliquoted to 1.5 ml microcentrifuge tubes and stored at -80°C.

TFB1 buffer

Rubidium chloride	100 mM
Manganese chloride	50 mM
Potassium acetate	30 mM

Calcium chloride	10 mM
Glycerol	15%

All the chemicals used were of molecular biology grade. pH (5.8) was adjusted carefully to avoid precipitation of insoluble manganese. After adjusting the pH, the buffer was filter sterilized and stored at 4°C.

TFB2 buffer

Morpholino Propanesulfonic acid (MOPS)	10 mM
Rubidium chloride	10 mM
Calcium chloride	75 mM
Glycerol	15%
pH	6.8

All the chemicals used were of molecular biology grade. pH was adjusted using KOH. After adjusting the pH, the buffer was filter sterilized and stored at 4°C.

3.6.4. Screening of transformants

Observed colonies on LB agar plates were subcultured and kept for incubation at 37°C overnight. PCR was performed to those subcultures to screen the positive clones. The analysis of transformants was done using gene specific primers. To determine the correct orientation of the gene cloned, a combination of vector specific primers and the gene specific primers were used *i.e.*, vector specific forward primer and gene specific reverse primer or the *vice versa.* Positive clones were inoculated to LB broth containing antibiotics (100 µg/ml ampicillin and 25 µg/ml kanamycin) and kept for incubation at 37°C for 5 hours. All the positive clones were

archived at -80°C in 30% glycerol broth and inoculated to LB butts to outsource for sequencing.

3.7. Sequencing of cloned genes

The positive clones having COI gene insert were sequenced to determine the orientation and correctness of nucleotide sequence using vector specific primers (Bangalore Genei, Bangalore) by outsourcing them to SciGenom Kerala. The sequences of COI gene of *Puntius filamentosus, Puntius assimilis, Barilius bakeri and Barilius canarensis* determined in this study have been submitted to GenBank.

The extent of homology between the nucleotides and the amino acid sequences was checked to known sequences in GenBank database using the respective BLAST programs (Altschul *et al.*, 1997) of the National Center for Biotechnology Information (NCBI, Bethesda, USA). The open reading frame (ORF) was checked using the ORF finder of the NCBI to check the correctness of ORF (http://www.ncbi.nlm.nih.gov/gorf/gorf.html).

3.8. Sequence analysis

The sequences were analyzed using different bioinformatics software. The extent of homology between the nucleotides and derived amino acid sequences to known sequences in GenBank was analyzed using the respective BLAST programs available at NCBI (http://www.ncbi.nlm.nih.gov/). Multiple alignments (Corpet, 1988) of the sequences was done to determine gaps, length and to display maximum similarity/conservation (http://blast.ncbi.nlm.nih.gov/bl2seq/wblast2.cgi). The open reading frame (ORF) was checked using the ORF finder to check the correctness of ORF (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The partial COI sequences of all the four fish were submitted to GenBank database (NCBI, Bethesda, USA).

3.8.1. Construction of Phylogenetic tree and analysis of genetic distance

A total of four sequences derived from the fishes under investigation (one each of *P.filamentosus*, *P. assimilis, B. bakeri* and *B. canarensis*) along with the other sequences of related species present in NCBI database were analyzed to construct a phylogenetic tree so as to check the position of a particular species in the history of life (Fig.22). The other sequences used were HM224200 (*Opsarius canarensis*), HM224197 (*O. bakeri*), and HQ654694 (*Channa striata*) where the last one i.e, *C. striata* was selected as an outgroup of the phylogram. Phylogenetic tree was constructed by using the software, MEGA5.0 (Tamura *et al.,* 2011).

A total of 6 sequences, i.e, four COI sequences of fishes under investigation along with the one each of *O. bakeri* and *O. canarensis* present in the NCBI database were used for the analysis of genetic distance. There was no COI sequence of either *P. filamentosus* or *P. assimilis* available in the NCBI database.

The genetic distances were calculated by using Kimura 2 parameter.

Results

IV. RESULTS

4.1. Morphometric comparison between *P. filamentosus* & *P. assimilis* and *B. canarensis* & *B. bakeri*

Morphometric evaluation is a key factor in taxonomical identifications through conventional approach. It is very important to identify the fish through Linnaean taxonomy and then proceed with molecular tool like DNA barcoding as a confirmative approach. The results of morphometric comparison between the fish species used in the present study are given in Table 4 and 5.

The main differences observed in *P. filamentosus & P. assimilis* were in case of maxillary barbel length, caudal peduncle length and dorsal to hypural distance, while the rest of the attributes almost remained same.

The main differences observed in *B. canarensis & B. bakeri* were in case of internarial-width, eye diameter and snout length, while the rest of the attributes showed no variation. It was not possible to distinguish between *P. filamentosus & P. assimilis* and *B. canarensis & B. bakeri* based on morphological characters alone as the differing characters were overlapping and insufficient to distinguish them.

4.2. Cloning of COI of *P. filamentosus*, *P. assimilis*, *B. canarensis* and *B. bakeri*

4.2.1. PCR amplification of COI

The total DNA extracted from the tissue sample of *Puntius filamentosus* was used as the DNA template for amplification of COI, yielded a PCR product of about 707 bp including the 52bp of primers used (Figure 8). Similarly, the amplification of COI gene of *P. assimilis*, *B. canarensis* and *B. bakeri* yielded an amplicon size of 655 bp (Fig: 11, 14 and 17).

Table 4. Morphometric characters of *P. filamentosus* and *P. assimilis*

Characters	*P. filamentosus* n = 5		*P. assimilis* n = 5	
	Range	Mean ± Stdv	Range	Mean ± Stdv
Standard length [mm]	47.3-83.3 mm		71.3-84.0mm	
% SL				
Head length	25.3-28.5	26.8±1.3	23.7-26.6	25.3±1.0
Head depth	17.7-20.1	18.7±1.1	18.2-20.4	19.4±0.9
Predorsal length	49.3-53.5	51.4±1.7	50.0-53.8	51.8±1.4
Dorsal to hypural distance	**52.2-55.8**	54.3±1.6	**57.0-58.5**	57.7±0.5
Maximum body depth	34.6-37.6	36.3±1.2	33.5-38.0	36.1±1.7
Maximum body width	12.4-21.0	15.6±3.9	15.3-17.3	16.4±0.9
Caudal peduncle length	**16.4-17.5**	17.0±0.5	**13.9-15.9**	15.0±0.8
Caudal peduncle depth	13.2-14.3	13.8±0.5	12.7-15.1	13.7±0.9
% HL				
Snout length	29.1-34.6	31.6±2.3	33.3-39.6	36.4±3.0
Eye diameter	32.7-39.7	35.6±3.0	32.1-37.6	34.7±2.5
Interorbital width	37.8-44.4	40.9±2.7	37.5-44.1	40.5±2.5
Internarial width	21.7-23.2	22.4±0.7	26.3-30.2	28.3±1.5
Maxillary barbel length	**6.6-8.6**	7.6±0.8	**11.5-20.6**	15.3±3.4
Postorbital head length	42.0-44.7	43.2±1.2	40.2-44.2	42.6±1.5

Table 5. Morphometric characters of *B. canarensis* and *B. bakeri*

Characters	*B. canarensis* n = 5		*B. bakeri* n = 5	
	Range	Mean ± Stdv	Range	Mean ± Stdv
Standard length [mm]	65.9-82.1 mm		42.3-68.4 mm	
% SL				
Head length	25.7-29.7	27.4±1.6	24.8-27.1	25.8±1.0
Head depth	17.7-23.4	20.5±2.0	18.8-20.8	19.5±0.8
Predorsal length	52.3-57.9	55.5±2.1	55.9-57.9	57.0±0.8
Dorsal to hypural distance	46.4-51.3	48.9±1.8	49.0-53.1	50.8±1.7
Maximum body depth	32.3-33.6	33.1±0.5	31.1-33.6	32.5±1.0
Maximum body width	13.0-14.2	13.6±0.4	12.0-14.4	13.2±0.9
Caudal peduncle length	16.4-19.1	17.8±1.0	14.3-18.8	16.3±1.6
Caudal peduncle depth	11.5-12.3	11.9±0.2	11.1-13.4	12.2±0.8
% HL				
Snout length	**26.4-33.6**	30.0±2.5	**22.4-24.7**	23.8±0.9
Eye diameter	**31.4-35.8**	34.2±1.8	**38.5-43.4**	40.8±2.0
Interorbital width	33.1-40.9	37.4±2.8	36.5-40.8	38.8±1.7
Internarial width	**20.6-25.1**	22.8±1.9	**18.2-19.8**	19.1±0.5
Postorbital head length	41.1-48.3	45.2±2.8	46.9-51.8	49.5±1.8

4.2.2. Cloning of gene coding for COI of *P. filamentosus*, *P. assimilis*, *B. canarensis* and *B. bakeri*

The purified PCR product of gene coding for COI of each species was ligated into pDrive cloning vector (Qiagen, Germany) and transformed to *E. coli* Ez competent cells. The transformed cells after incubation were spread plated on LB agar plates containing antibiotics (100 µg/ml ampicillin and 25 µg/ml kanamycin). The colonies formed after incubation was tested for the presence of inserts through PCR using gene specific primers. The positive clones containing COI gene inserts of all four fish species showed a PCR product of 707 bp including the 52bp of primers used (Figures 9, 12, 15 and 18).

4.2.3. Sequencing of recombinant clones

Sequencing of the PCR confirmed positive clones of COI gene of each species was done to check the orientation, nucleotide sequences and amino acid sequences. The analysis of the cloned products revealed that the gene was of partial length and with correct reading frame and correct translation was obtained. The nucleotide and deduced amino acid sequences of COI gene of all four fish species is shown in Fig. 10, 13, 16 and 19.

4.3. Sequence analysis

A total of four partial COI sequences, one each from *P. filamentosus*, *Puntius assimilis, B. bakeri* and *B. canarensis* of 655 bp length with no insertions, deletions or stop codons were obtained in this study and have been submitted to Genbank under the accession numbers JN255695, JN255696, JN255697 and JN255698.

4.3.1. Comparison of partial COI sequences of *P. filamentosus* and *P. assimilis*

In order to check the sequence similarities and differences between *P. filamentosus* and *P. assimilis*, comparison was carried out between the derived partial sequences of the gene COI

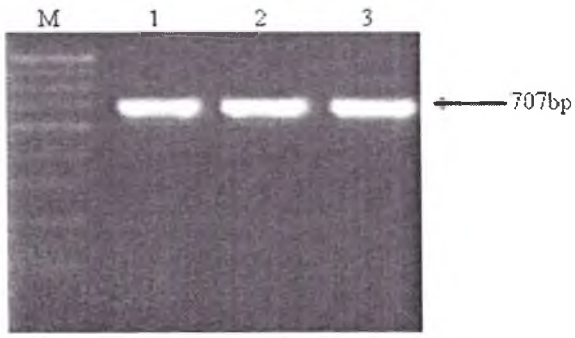

Fig.8. PCR amplification and detection of a 655 bp region of COI of *Puntius filamentosus* by gel electrophoresis

Lane M: 100 bp DNA Ladder

Lanes 1-3: (COI gene of 655bp+ 52 bp of primer) of *Puntius filamentosus*

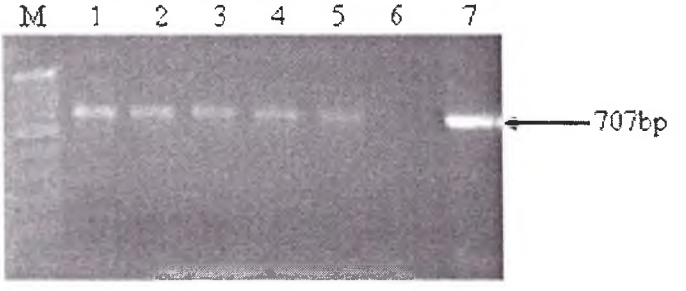

Fig.9. PCR amplification of recombinant COI clones of *Puntius filamentosus* by gene specific primers.

Lane M: 100 bp DNA Ladder

Lanes 1, 2, 3, 4, 5: Positive (COI of 655bp+52 bp of primers) clones

Lane 6: Negative control

Lane 7: Positive control

```
   1
   1  cctttatctagtatttggtgcctgagccggaatagtaggaaccgccttaagcctccttatc

  21
  61  cgggctgaactaagtcagccagaatcactattaggcgacgaccaaatttataacgttatc

  41
 121  gttactgcccacgcctttgtaataattttctttatagtaataccaatccttattggagga

  61
 181  tttgggaactgactagtgccactaataattggagctcctgacatagcattcccacgaata

  81
 241  aacaacataagcttctgactattgcccccatcattcctactcctactagcctcctctggt

 101
 301  gtcgaagctggggcaggaacaggatgaacagtatatccacccccttgcaggaaacctagcc

 121
 361  cacgccggagcatccgttgacctaacaattttctcactccacttagcaggtgtatcatca

 141
 421  atccttggagcaattaattttatcaccacaacaattaacataaaaccccccaacgatctca

 161
 481  caatatcaaacaccactgtttgtctgatctgtacttgtaaccgctgtattactattactc

 181
 541  tcactaccagtattagctgccggaatcacaatactcctaacagatcgaaatcttaacacc

 201
 601  acattctttgacccagcaggtggaggagacccaatcctgtaccaacacctattc
```

Fig.10. Partial nucleotide sequence and deduced amino acid sequence (Start- second nucleotide) of COI of *Puntius filamentosus*.

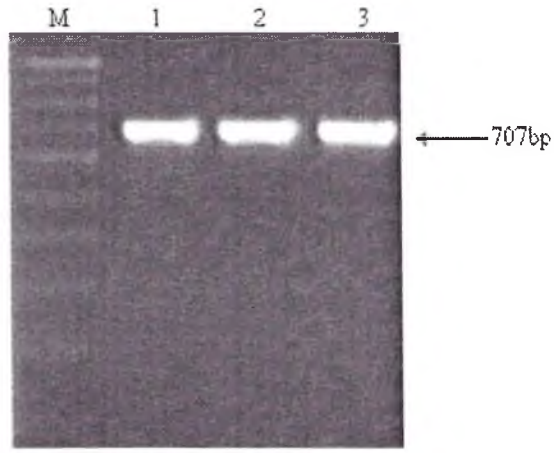

Fig.11. PCR amplification and detection of a 655 bp region of COI of *Puntius assimilis* by gel electrophoresis

Lane 1: 100 bp DNA Ladder

Lanes 1-3: (COI gene of 655bp+ 52 bp of primer) of *Puntius assimilis*

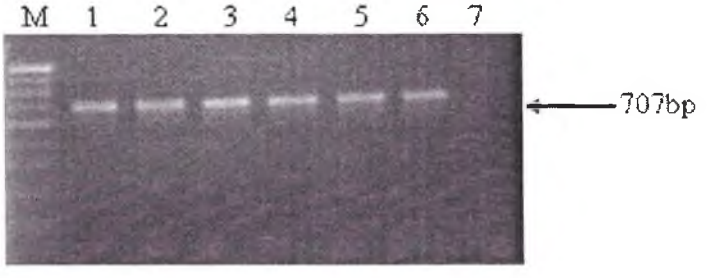

Fig.12. PCR amplification of recombinant COI clones of *Puntius assimilis* by gene specific primers.

Lane M: 100 bp DNA Ladder

Lanes 1, 2, 3, 4, 5: Positive (COI gene of 655bp+52 bp of primers) clones

Lane 6: Positive control

Lane 7: Negative control

```
   1
   1  cctatacctgatatttggtgcctgagccgggatagtaggaactgccctaagcctcctaatc

  21
  61  cgagctgaactaagccaaccaggaacacttctaggtgatgatcaaatctataatgtcatc

  41
 121  gtcactgcccacgcctttgtaataatcttctttatagtaataccaatcctcatcggagga

  61
 181  ttcgggaactggctagtaccattaataattggggcacctgatatagcattcccgcgaatg

  81
 241  aacaatataagcttttgacttctgccccccatcattcttactactactaacctcctctggc

 101
 301  gttgaggcaggggccggcactgggtggacagtatatccgccccttgcaggaaatttagcc

 121
 361  cacgctggggcatcagttgacctgacaatttttttcactccatctagcaggagtctcatcg

 141
 421  atcctaggggcaatcaattttattaccacaattattaacataaaaccccccaaccatctca

 161
 481  caataccagacaccactattcgtctgatctgtcctagtaaccgccgttctattattactc

 181
 541  tcactaccaatcctagccgctggaattacaatgctcctgacagatcgaaaccttaacaca

 201
 601  acattctttgacccagcaggcgggggagaccctatcctataccaacacctatttt
```

Fig.13. Partial nucleotide sequence and deduced amino acid sequence (Start- second nucleotide) of COI of *Puntius assimilis*

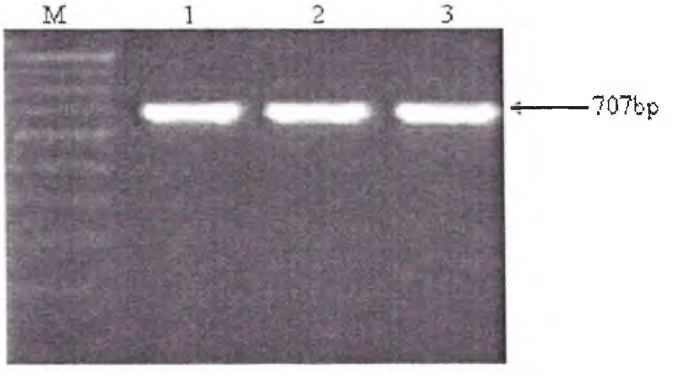

Fig.14. PCR amplification and detection of a 655 bp region of COI of *Barilius canarensis* by gel electrophoresis

Lane M: 100 bp DNA Ladder

Lanes 1-3: (COI gene of 655bp+ 52 bp of primer) of *Barilius canarensis*

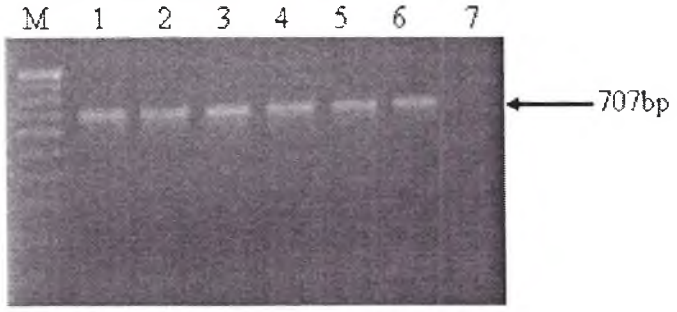

Fig.15. PCR amplification of recombinant COI clones of *Barilius canarensis* by gene specific primers.

Lane M: 100 bp DNA Ladder

Lanes 1, 2, 3, 4, 5: Positive (COI of 655bp+52 bp of primers) clones

Lane 6: Positive control

Lane 7: Negative control

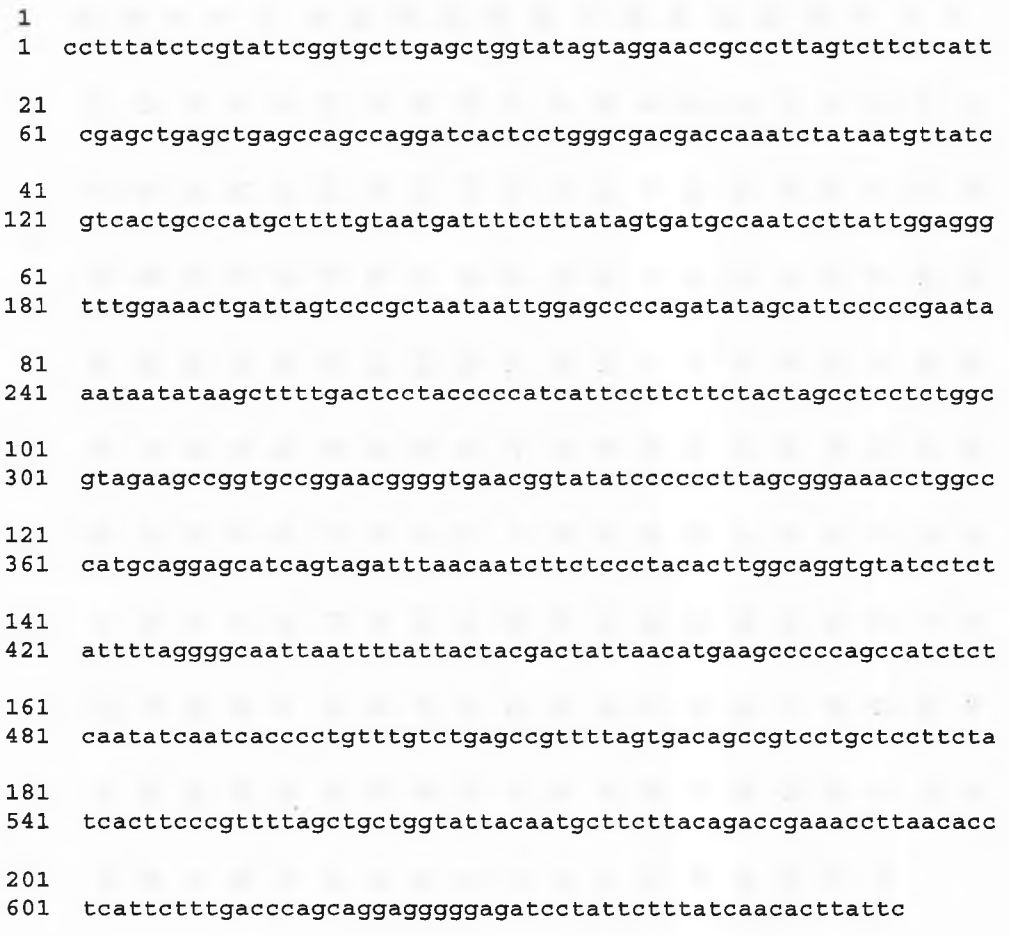

Fig.16. Partial nucleotide sequence and deduced amino acid sequence (Start- second nucleotide) of COI of *Barilius canarensis.*

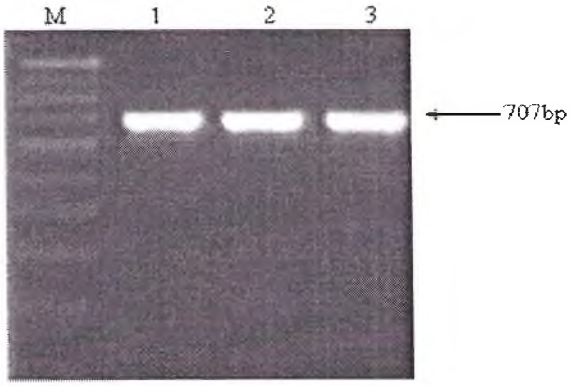

Fig.17. PCR amplification and detection of a 655 bp region of COI of *Barilius bakeri* by gel electrophoresis

Lane M: 100 bp DNA Ladder

Lanes 1-3: (COI gene of 655bp+ 52 bp of primer) of *Barilius bakeri*

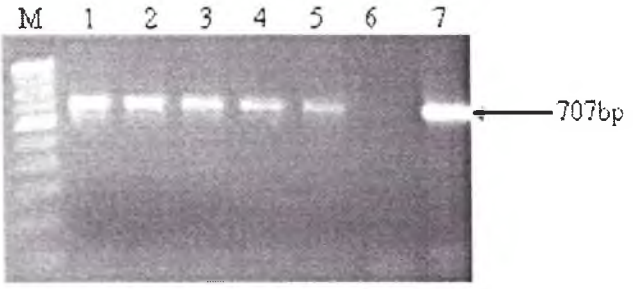

Fig.18. PCR amplification of recombinant COI clones of *Barilius bakeri* by gene specific primers.

Lane M: 100 bp DNA Ladder

Lanes 1, 2, 3, 4, 5: Positive (COI of 655bp+52 bp of primers) clones

Lane 6: Negative control

Lane 7: Positive control

```
   1
   1  cctttatctcgtattcggtgcttgagctggtatagtaggaaccgcccttagccttctcatt

  21
  61  cgagctgagctaagccagccaggatcactcctgggcgacgaccaaatttataatgttatt

  41
 121  gttactgcccatgcttttgtaatgattttctttatagtgatgccaatccttattgggggg

  61
 181  tttggaaactgactagtcccactaatgattggagccccagatatagcattcccccgaata

  81
 241  aacaatataagcttttgactcctacccccgtcattccttcttcttctagcctcctctggt

 101
 301  gtagaagccggtgccggaacggggtgaacggtatatccccccttagcgggaaatctggct

 121
 361  catgcaggagcatcggtggatttaacaatcttctccctacacttggcaggtgtatcctcc

 141
 421  attttaggggcaattaattttattactacaacgattaacataaaaccccccagccatctcc

 161
 481  cagtatcaaacacccctatttgtctgagccgttttagtgacggccgtcctactccttcta

 181
 541  tcacttcccgttttagctgctggtattacaatgcttcttacagaccgaaaccttaatacc

 201
 601  tcattctttgatccagcaggaggggggagatcctattctttatcaacacttattt
```

Fig.19. Partial nucleotide sequence and deduced amino acid sequence (Start- second nucleotide) of COI of *Barilius bakeri.*

by using the program multalin (Corpet, 1988). The result is shown in the Fig.20. The resulted picture clearly shows that the two of fishes being compared are entirely different as they show differences in 63 nucleotides (9.61%).

4.3.2. Comparison of partial COI sequences of *Barilius canarensis* and *Barilius bakeri*

In order to check the sequence similarities and differences between *B. canarensis* and *B. bakeri*, comparison was carried out between the derived partial sequences of the gene COI by using multalin (Corpet, 1988). The result is shown in the Fig. 21.

The resulted picture clearly shows that the two of the fishes being compared are entirely different as they show differences in 31 nucleotides (4.88%).

To check the sequence differences between *P. filamentosus*, *P. assimilis B. canarensis* and *B. bakeri*, comparison was carried out between the derived partial sequences of the gene COI by using multalin (Corpet, 1988). The result is shown in the Fig. 22.

4.3.3. Genetic distance, phylogram and analysis at inter and intraspecific levels of the sequences of *Puntius filamentosus*, *Puntius assimilis*, *Barilius canarensis* and *Barilius bakeri*

The result clearly shows the interspecific distance between *P. filamentosus* and *P. assimilis*, *B. bakeri* and *B. canarensis*. Distribution of K2P distance (percent) for COI within the four fish barcoded and the others derived from NCBI database is given in the Table 6.

The phylogram (Fig. 23) generated through neighbour joining method using K2P distance was highly reliable as, outgroup used (*Channa striata*) was clearly segregated in a different clade as anticipated. The ability of COI to differentiate two genera was evident through the phylogram as the two clades (Clade A and Clade B) were distinguishable.

The generated phylogenetic tree with 1000 bootstrap replications clearly showed that the two species *P. filamentosus* and *P. assimilis* were entirely different with a genetic distance of 16.4% and were positioned at different regions with a bootstrap value of 99. Similarly, *B. canarensis* and *B. bakeri* could be clearly differentiated with a genetic distance of 14.6% and with a bootstrap value of 100. The average genetic distance between *Puntius* and *Barilius* was 14.72%.

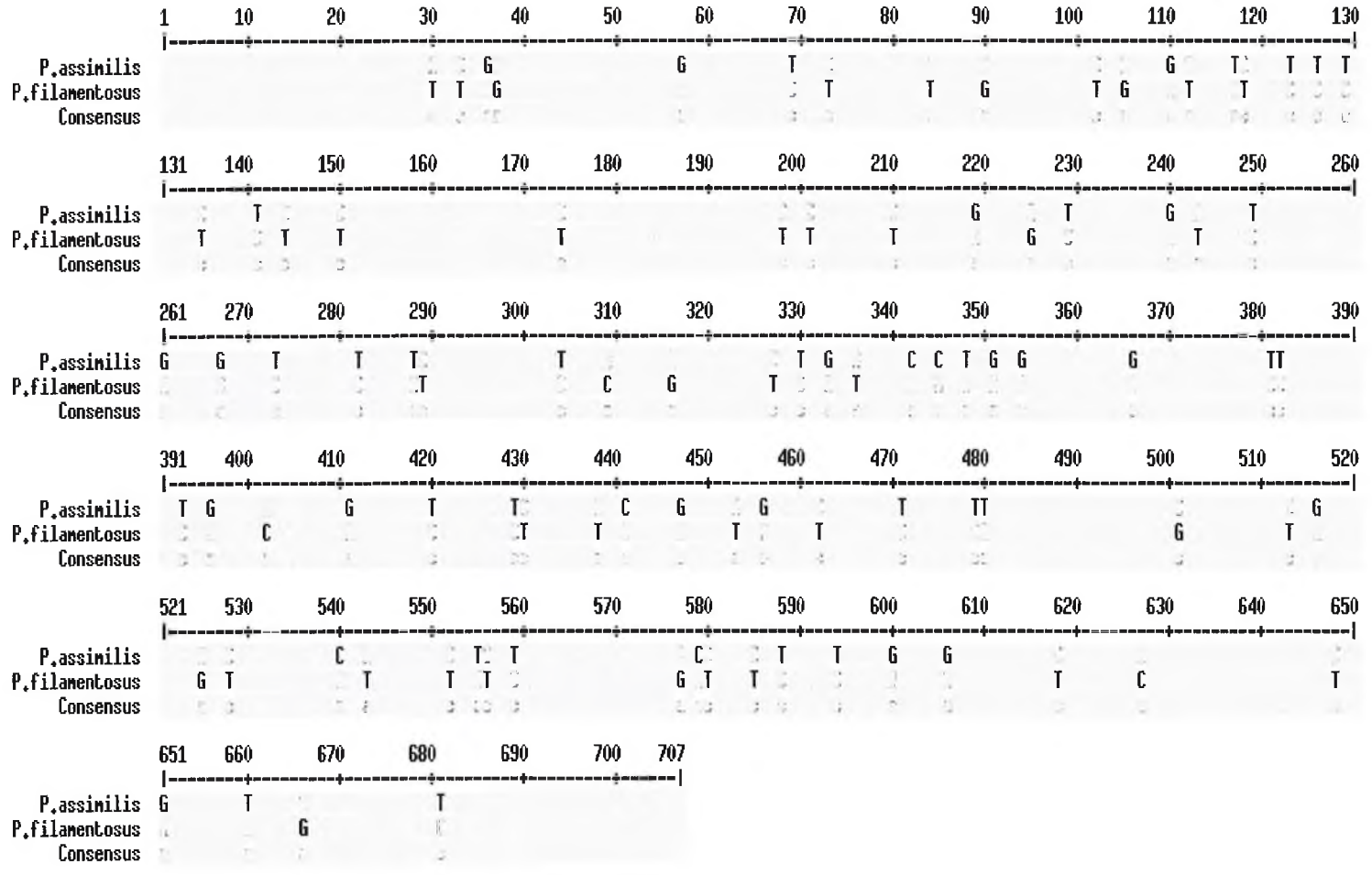

Fig.20. Comparison of COI gene sequences of *Puntius filamentosus* and *Puntius assimilis* by multalin.

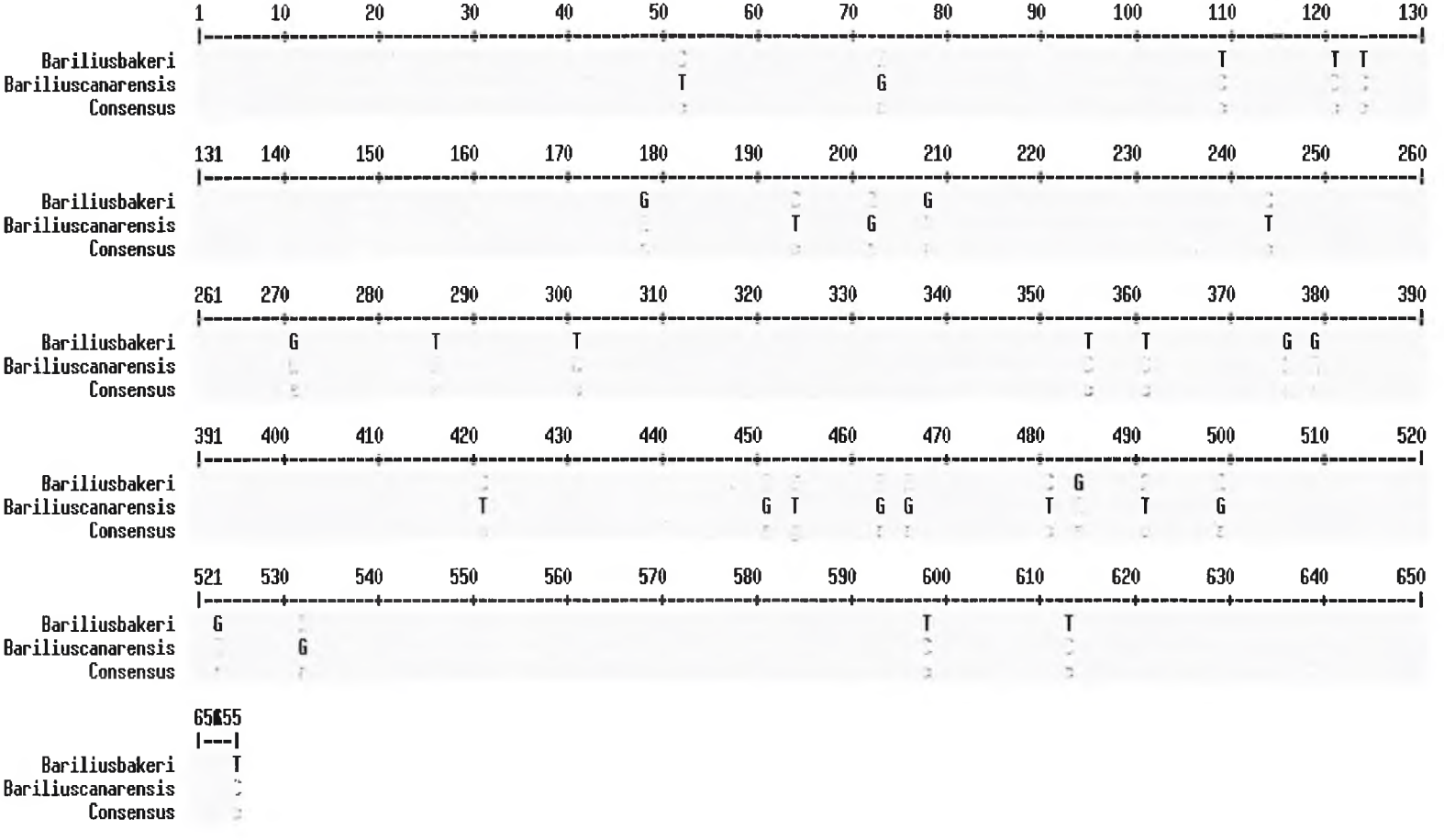

Fig.21. Comparison of COI sequences of *Barilius canarensis* and *Barilius bakeri* by multalin.

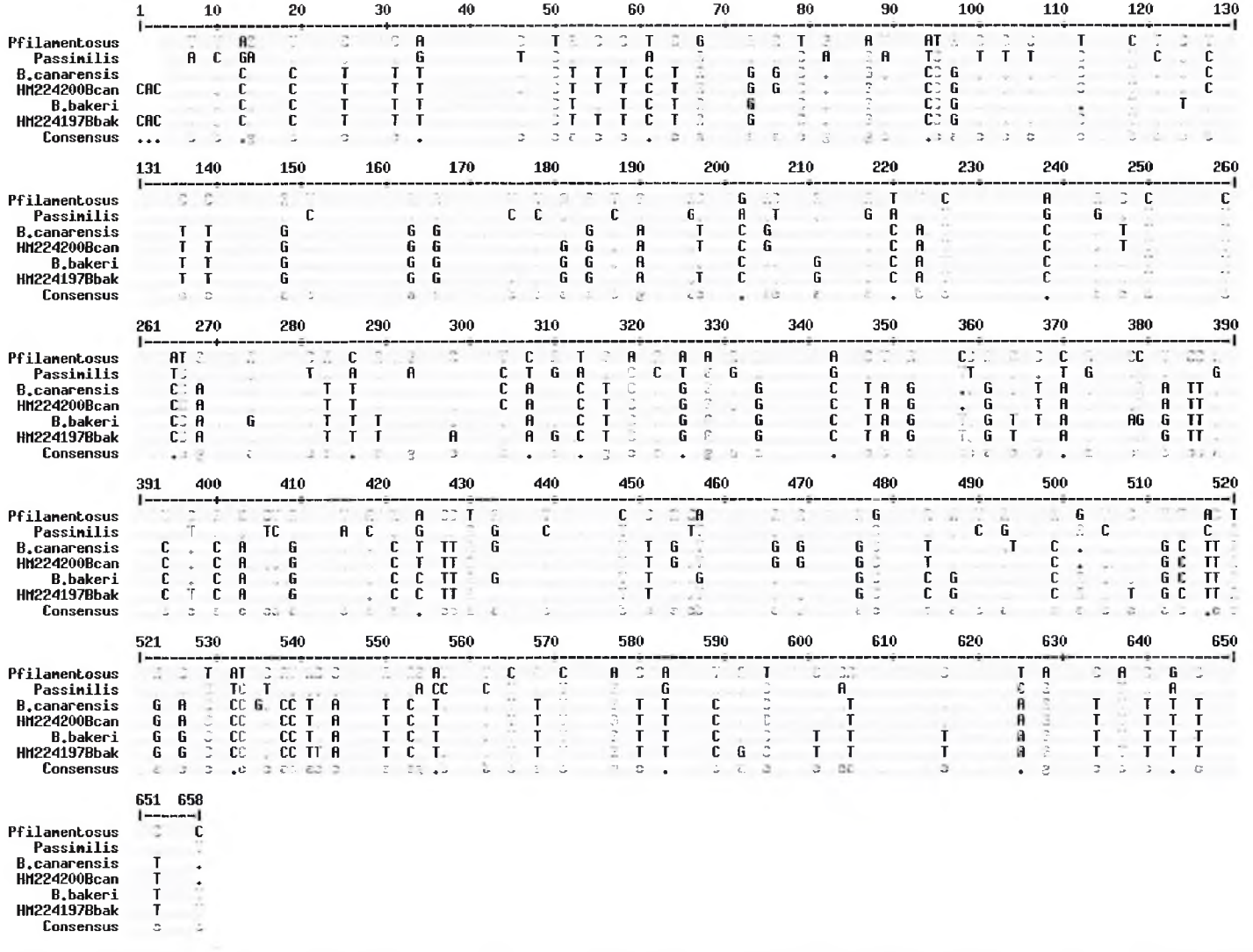

Fig. 22. Comparison between the six fish used in the phylogram, showing the difference between their nucleotides.

Table.6. Distribution of K2P distance (percent) for COI within the four fishes barcoded and the others derived from NCBI database.

Fishes	1	2	3	4	5
**P. filamentosus* (1)					
**P. assimilis* (2)	16.4				
B. canarensis HM224200(3)	20.6	24.2			
**B. canarensis* (4)	21.1	24.2	0.6		
*B. bakeri*HM224197 (5)	21.9	24.2	4.4	5.1	
**B. bakeri* (6)	21.1	24.6	4.8	5.1	2.5

*Specimens used in the present study.

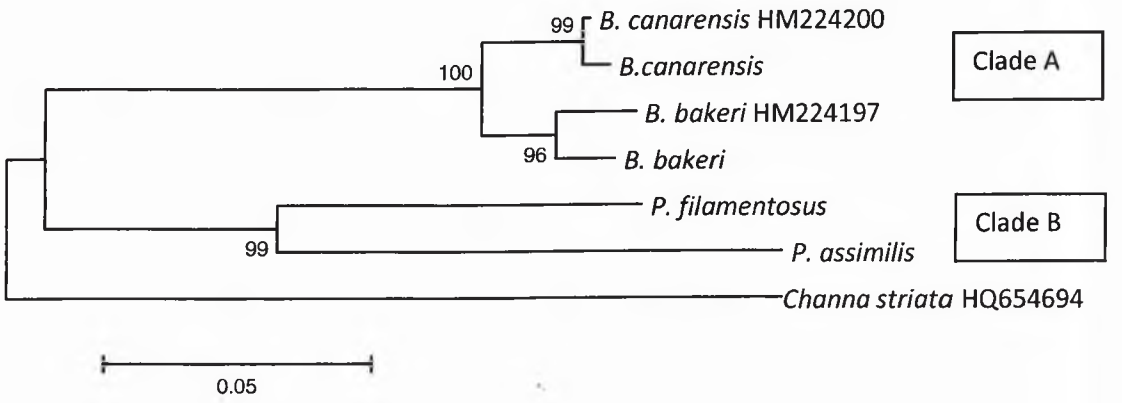

Fig.23. Neighbour-joining tree (with bootstrap values; 1000 replicates) of COI sequences of fishes constructed using K2P distance. *P. filamentosus*, *P. assimilis*, *B. bakeri* and *B. canarensis* correspond to the fishes used in this study. Scale bar: 0.05 substitutions per site. (Note: *Opsarius* synonym *Barilius*).

Discussion

V. DISCUSSION

Fishes from the genus *Puntius* and *Barilius* play notable role in the world ornamental fisheries as these fishes are colorful and have good value in the market. As the species from each genus do not show much phenotypical difference with each other, it became a problem to identify them. Conventional taxonomy which is based on morphological characters failed to distinguish between these closely related species.

Although all the attributes of morphometric analysis remained almost same, a major difference could be observed between *P. assimilis* from *P. filamentosus* in three characters. For *P. filamentosus*, the dorsal to hypural distance was 52.2-55.8% of standard length, caudal peduncle length was 16.4-17.5% of standard length and maxillary barbel length was 6.6-8.6% of head length. The corresponding values observed for *P. assimilis* was 57.0-58.5% of SL, 13.9-15.9% of SL and 11.5-20.6% of HL, respectively.

Further, in case of *B. bakeri* and *B. canarensis* also we could observe the major difference in the three attributes though the other attributes remained almost same. For *B. canarensis*, snout length was 26.4-33.6% of head length, eye diameter was 31.4-35.8% of HL and internarial width was 20.6-25.1% of HL. The corresponding values obtained for *B. bakeri* were 22.4-24.7 % of HL, 38.5-43.4% of HL and 18.2-19.8% of HL, respectively. These differences in the comparative attributes could give us a hint that they were not same but could not precisely differentiate the species which have been compared.

The difficulty in differentiating closely resembling fish species based on morphological characters leads to misidentification, which could have serious consequences especially in ornamental fish trade. In such situations genetic information is expected to aid in establishing species identity. In the ornamental fish industry, *P. assimilis* is traded as *P. filamentosus* and

B. canarensis as *B. bakeri* because of their close similarity. In situations like this, the molecular based approach involving nucleotide sequencing is expected to resolve the problem. Among many genetic markers, the mitochondrial gene, COI, is found to be unique and species specific. Using the sequence data of this gene, it is possible to establish species specific signature sequence. This approach called DNA barcoding is capable of differentiating all life forms based on COI sequence (Hebert, 2003a). Barcoding is very useful in biodiversity studies as it helps not only faster identification of species, but also in identifying eggs, larvae or fragments of a species which could not be identified by morphology based taxonomic approach. In this regard, the initiative at international level to barcode all life forms has been successful in establishing barcode of life database (BOLD), and for fishes, FISHBOL.

The sequences generated by targeting COI were analyzed so as to differentiate these fishes based on the genetic distance between them. As the species from each genus do not show much phenotypical difference with each other, it became a problem to identify them using Linnaean taxonomy. Conventional taxonomy which is based on morphological characters failed to distinguish between these closely related species. DNA barcoding combined with the conventional taxonomy solved the confusion in identifying and differentiating *P. assimilis* from *P. filamentosus* as well as *B. canarensis* from *B. bakeri*.

The sequences generated by targeting COI were analyzed so as to differentiate these fishes based on the genetic distance between them. In the studies conducted earlier by the usage of K2P genetic distance for analyzing the data exposed that barcode variation keeps on increasing from species to genera and so on (Avise *et al.,* 1987, Black *et al.,* 1997, Blaxter and Floyd 2003, Borda and Siddal, 2004, Brower 1999, Hebert *et al.,* 2003, Khan *et al.,* 2010). But this was not observed to be true in this study. For example; the genetic distance between *B.*

canarensis & *Barilius bakeri* was 14.6% and that of *P. filamentosus* and *P. assimilis* was 16.4%. But genetic distance between the genera *Puntius* and *Barilius* was 14.72%.

Though these two pair of fish look- alike morphologically, they were observed to be distant organisms molecularly. The reason for the high intra-specific genetic distance between *P. filamentosus* and *P. assimilis* was hypothesized that the species which was showing all the characters of being *P. assimilis* might have evolved in a faster rate to that of *P. filamentosus*, leading to higher genetic distance. However, this study proves the efficiency of partial sequence of COI in identifying species. The ability of COI in differentiating the two species could be evident from both the clades A and B. The bootstrap values of all clades were good enough without any exceptions. On the other hand it may not be possible to extract the true phylogeny of fishes from only a 655 bp COI through K2P distance and neighbor joining as COI does not determine the entire phenotypic characters. More gene regions should be used including nuclear genes (Ward *et al.,* 2005) and few more gene regions such as Cyt b, and 16s RNA of mitochondria and it is also recommended to combine techniques of DNA barcoding and microarrays (Kochzius *et al.*,2010).

The use of both nuclear and mitochondrial sequences as the best strategy to recover relationships that most likely reflect the phylogenetic history of these lineages is recommended (Paton and Baker, 2006). Nevertheless, barcoding discriminates all the fishes under study and would clearly be capable of identifying individually examined fins, eggs, larvae and the young ones of these fishes. Bootstrap values of all the fishes were on higher side as the minimum of 47 was observed on clade B. This study has strongly validated the efficiency of COI barcodes to identify *P. filamentosus, P. assimilis, B. canarensis* and *B. bakeri* from their type locality.

VI. SUMMARY

A promising way for the taxonomical identification of fish is the use of DNA barcoding. DNA barcoding can be done by using the markers such as Cytochrome c oxidase subunit 1 (COI), Cytochrome b, 16s RNA etc. However, DNA barcoding by the use of partial sequencing of COI as a marker is widely practiced. The present work was an attempt to give the correct identity to the fishes; *Puntius filamentosus*, *Puntius assimilis*, *Barilius canarensis* and *Barilius bakeri* by supporting the conventional taxonomy with DNA barcoding data as these fishes had identity crisis.

In the present study the COI genes were amplified using extracted genomic DNA by PCR, cloned in pDrive vector and sequenced. The isolated COI gene sequences of all four fishes were deposited in GenBank database under the accession numbers JN255695, JN255696, JN255697 and JN255698. The partial COI sequences of *P. filamentosus* & *P. assimilis* and *B. canarensis* & *B.bakeri* were aligned and compared for the similarities and differences in the nucleotide sequence. The genetic distance between *P. filamentosus* and *P. assimilis* was 16.4% and that of *B.bakeri* and *B. canarensis* was 14.6%. which clearly showed that each of the fishes are different though they look alike phenotipically (except the mouth region and coloration in case of *P.filamentosus* and *P. assimilis* and the band pattern on the body surface in case of *B. canarensis* and *B.bakeri*). The constructed phylogenetic tree could reveal that these fishes fit in different positions in the tree of life.

The present work could revalidate the identity of the two fishes used. *P. assimilis* is entirely different fish and it should not be confused with *P. filamentosus*. Similarly, *B. canarensis* should not be confused with the variant of *B. bakeri*.

VII. BIBLIOGRAPHY

ALLANDER, T., EMERSON, S. E., ENGLE, R. E., PURCELL, R. H. and BUKH, J. 2001. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc. Natl. Acad. Sci. USA*, **98**(20):11609-11614

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. and LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**:3389-3402.

AVISE, J. C., ARNOLD, J., BALL, R. M., BERMINGHAM, E., LAMB, T., NEIGEL, J. E., REEB, C. A. and SAUNDERS, N. C. 1987. Intraspecific phylogeography: The mitochondrial bridge between population genetics and systematics. *Annu. Rev. Ecol. Evol. S.*, **18**: 489–522

AVISE, J. C. 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution*, **43**(6): 1192–1208

AYALA, F. J. 1983. Enzymes as taxonomic characters. In : Protein Polymorphism: Adaptive and Taxonomic Significance (G. S. Oxford and D. Rollinson, Eds.), Systematics Association Special **24**: 3–26, Academic Press, London

BALLARD, J. W. O. and KREITMAN, M. 1995. Is mitochondrial DNA a strictly neutral marker? *Tree*, **10**: 485–488

BARRETT, R. D. H. and HEBERT, P. D. N. 2005. Identifying spiders through DNA barcodes. *Can. J. Zool.*, **83**: 481-491

BENSASSON, D., ZHANG, D. X., HARTI, D. L. and HEWITT, G. M. 2001. Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends. Ecol. Evol.*, **16**(6): 314-321

BIRSTEIN, V. J., DESALLE, R., DOUKAKIS, P., HANNER, R., RUBAN, J. I. and WONG, E. 2009. Testing taxonomic boundaries and the limit of DNA barcoding in theSiberian sturgeon, *Acipenser baerii*. *Mt. DNA*, **20** (5–6): 110–118

BLACK, M. B., HALANYCH, K. M., MAAS, P. A. Y., HOEH, W. R. and HASHIMOTO, J. 1997. Molecular systematic of vestimentiferan tubeworms from hydrothermal vents and cold water seeps. *Mar. Biol.*, **130**: 141-149

BLAXTER, M. 2004. The promise of DNA taxonomy. *Phil. Trans. R. Soc. Lond.* B., **359**: 669-679

BLAXTER, M. and FLOYD, R. 2003. Molecular taxonomics for biodiversity surveys: Already a reality. *Trends. Ecol. Evol.*, **18**: 268-269

BLAXTER, M., MANN, J., CHAPMAN. T., THOMAS, F., WHITTON, C., FLOYD, R. and ABEBE, E. 2005. Defining operational taxonomic units using DNA barcode data. *Phil. Trans. R. Soc. Lond.* B., **360**(1462): 1935-1943

BORDA, E. and SIDDAL, M. E. 2004. Arhynchobdellida (Annelida: Oligochaeta: Hirunida): Phylogenetic relationships and evolution. *Mol. Phylogenet. Evol.*, **30**: 213-225

BRITZ, R., CONWAY, K. W. and RUBER, L. 2009. Spectacular morphological novelty in a miniature cyprinid fish, *Danionella dracula*. *Proc. R. Soc. Lond.* B., **276**: 2179-2186

BROWER, A. V. Z. 1999. A crytique of Davis and Nixon's population aggregation analysis. *Syst. Biol.*, **48**: 199-213

BROWN, W. M., GEORGE, M. JR. and WILSON, A. C. 1979. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*, **76**(4): 1967-1971

CHASE, J. M. 2005. Towards a really unified theory for metacommunities. *Funct Ecol,* **19**: 182–186

CIVERA, T. 2003. Species identification and safety of fish products. *Vet. Res. Commun.*, **27**:481–489

CORPET, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.*, **16** (22), 10881-10890

COX, A. J. and HEBERT, P. D. N. 2001. Colonization, extinction, and phylogeographic patterning in a freshwater crustacean. *Mol. Ecol.*, **10** (2): 371-386

DALEBOUT, M. L., BAKER, C. S., MEAD, J. G., COCKCROFT, V. G. and YAMADA, T. K. 2004. A comprehensive and validated molecular taxonomy of beaked whales, family Ziphiidae. *J. Hered.*, **95** (6): 459–473

DAY, F. 1878. The Fishes of India: Being a natural history of the fishes known to inhabit the seas and freshwaters of India, Burma and Ceylon. William Dowson & Co., London, 778pp

DOYLE, J. J. and GAUT, B. S. 2000. Evolution of genes and taxa: a primer. *Plant. Mol. Biol.*, **42**: 1–23

DUNN, C. P. 2003. Keeping taxonomy based in morphology. *Tree*, **18**: 270–271.

FERGUSON, J. W. H. 2002. On the use of genetic divergence for identifying species. *Biol. J. Linn. Soc.*, **75**: 509-516

FLOYD, R., ABEBE, E., PAPERT, A. and BLAXTER, M. 2002. Molecular barcodes for soil nematode identification. *Mol. Ecol.*, **11**: 839-850

FOLMER, O., BLACK, M., HOETH, W., LUTZ, R. and VRIJENHOEK, R. 1994. DNA primers for amplification of mithocondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.*, **3**(5): 294-299

GODFRAY, H. C. J. 2002. Challenges for taxonomy- the discipline will have to reinvent itself if it is to survive and flourish. *Nature*, **417**: 17–19

HAJIBABAEI, M., SMITH, M. A., JANZEN, D. H., RODRIGUEZ, J. J., WHITFIELD, J. B. and HEBERT, P. D. N. 2006c. A minimalist barcode can identify a specimen whose DNA is degraded. *Mol. Ecol. Notes*, **6**: 959-964

HAMILTON- BUCHANAN, F. 1822. An Account of the fishes of river Ganges and its branches. George Ramsay and Co., London, vii+405pp, 39 pls

HAWKSWORTH, D. L. and KALIN-ARROYO, M. T. 1995. Magnitude and distribution of biodiversity. In: Heywood VH and Watson RT (Eds) *Global biodiversity assessment, pp.* 107-191

HEBERT, P. D. N., CYWINSKA, A., BALL, S. L. and DEWAARD, JR. 2003a. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond.* B., **270**: 313-321

HEBERT, N., HANNER, R., HOLM, E., MANDRAK, N. E., TAYLOR, E., BURRIDGE, M., WATKINSON, D., DUMONT, P., CURRY, A., BENTZEN, P., ZHANG, J., APRIL, J. and BERNATCHEZ, L. 2008. Identifying Canadian freshwater fishes through DNA barcodes. *PLoS ONE*, **3** (6): e2490

HEBERT, P. D. N., PENTON, E. H., BURNS, J. M., JANZEN, D. H. and HALLWACHS, W. 2004b. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator. Proc. Natl. Acad. Sci. U S A*, **101**: 14812-14817

HEBERT, P. D. N., RATNESINGHAM, S. and DEWAARD, JR. 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. Biol. Sci.*, **270**: 96- 99

HEBERT, N., TROTTIN, E. D., IRISSON, J. O., MEYER, C. and PLANES, S. 2010. Identifying coral reef fish larvae through DNA barcoding: A test case with thefamilies Acanthuridae and Holocentridae. *Mol. Phylogenet. Evol.*, **55**: 1195–1203

JAYARAM, K. C. 1999. The Freshwater Fishes of Indian Region. Narendra Publishing House, New Delhi, 455pp

JERDON, T. C. 1848–49. On the freshwater fishes of southern India. *Madras Journal of Literature and Science*, **15** (1848): 141–149; **15** (1849): 302–346

JUAN, C., IBRAHIM, K. M., OROMÌ, P. and HEWITT, G. M. 1996. MitochondrialDNA sequence variation and phylogeography of *Pimelia* darkling beetles on the island of Tenerife (Canary Islands). *J. Heredity*, **77**: 589-598

KARTAVTSEV, Y. P., JUNG, S., LEEA, Y., BYEON, H. and LEEA, J. 2007. Complete mitochondrial genome of the bullhead torrent catfish, *Liobagrusobesus* (Siluriformes, Amblycipididae): Genome description and phylogenetic considerations inferred from the Cyt b and 16S rRNA genes. *Gene*, **396**: 13–27

KARTAVTSEV, Y. P., SHARINA, S. N., GOTO, T., CHICHVARKHIN, A., BALANOV, A., VINNIKOV, K. A., IVANKO, V. N and HANZAWA, N. 2008. Cytochrome oxidase 1 gene sequence analysis in six flatfish species (Teleostei, Pleuronectidae) of Far East Russia with inferencesin phylogeny and taxonomy. *Mt. DNA*, **19**(6): 479–489

KHAN, S. A., LYLA, P. S., JOHN, B. A., KUMAR, C. P., MURUGAN, S. and JALAL, K. C. A. 2010. DNA barcoding of *Stolephorus indicus*, *Stolephorus commersonii* and *Terapon jarbua* of Parngipettai Coastal Waters. *Biotechnology*, **9** (3): 373-377

KNOWLTON, N. 1993. Sibling species in the sea. *Annu. Rev. Ecol. Evol. S.*, **24**: 189-216

KOCHZIUS, M., SEIDEL, C., ANTONIOU, A., BOTLA, SK. and CAMPO, D., et al. 2010. Identifying Fishes through DNA Barcodes and Microarrays. *PLoS ONE*, **5**(9): e12620

KORPELAINEN, H. 2004. The evolutionary processes of mitochondrial and chloroplast genomes differ from those of nuclear genomes. *Naturwissenschaften*, **91**: 505–518

KRESS, W. J., WURDACK, K. J., ZIMMER, E. A., WEIGT, L. A. and JANZEN, D. H. 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U S A*, **102**: 8369–8374

LINNAEUS, C. 1756. Systema Naturae. Ninth edition. Theodor Haak, Leiden (Lugdunum Batavorum)

LAKRA, W. S., VERMA, M. S., GOSWAMI, M., LAL, K. K., MOHINDRA, V., PUNIA P., GOPALAKRISHNAN, A., SINGH, K. V., WARD, R. D., and HEBERT, P. 2009. DNA barcoding Indian marine fishes. *Mol. Ecol. Resour.*, doi:10.1111/j.1755-0998.2010.02894.x

LIPSCOMB, D., PLATNICK, N. and WHEELER, Q. 2003. The intellectual content of taxonomy: a comment on DNA in taxonomy. *Trends. Ecol. Evol.*, **18**: 65–67

LOWENSTEIN, J. H., AMATO, G. and KOLOKOTRONIS, S. O. 2009.The real maccoyii: Identifying tuna sushi with DNA barcodes – contrasting characteristic attributes and genetic distances. *PLoS ONE*, **4** (11): e7866 doi:10.1371/journal.Pone.0007866

LYNCH, M. and JARRELL, P. E. 1993. A method for calibrating molecular clocks and its application to animal mitochondrial DNA. *Genetics*, **135**: 1197-1208

MANWELL, C. and BAKER, C. M. A. 1963. A sibling species of sea-cucumber discovered by starchgel electrophoresis. *Comp. Biochem. Physiol.*, **10**: 39–53

MARKMANN, M. and TAUZ, D. 2005. Reverse taxonomy: an approach towards determining the diversity of meiobenthic organisms based on ribosomal RNA signature sequences. *Phil. Trans. R. Soc. B.*, **360** (1462): 1917-1924

NEWMASTER, S. G., FAZEKAS, A. J. and RAGUPATHY, S. 2006. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. *Can. J. Bot.*, **84**: *335-341*

OLMSTEAD, R. G. and PALMER, J. D. 1994. Chloroplast DNA systematics: A review of methods and data analysis. *Am. J. Bot.*, **81**(9): 1205-1224

NANNEY, D. L. 1982. Genes and phenes in *Tetrahymena. Bioscience*, **32**:783–788.

NEIGEL, J., DOMINGO, A. and STAKE, J. 2007. DNA barcoding as a tool for coral reef conservation. *Coral Reefs*, **26**(3): 487-499

PACE, N. R. 1997. A molecular view of microbial diversity and the biosphere. *Science*, **276** (5313): 734-740

PADIAL, J. M. and DE LA RIVA, I. 2009. Integrative taxonomy reveals cryptic Amazonian species of *Pristimantis* (Anura). *Zool. J. Linn. Soc.*, **155**: 97–122

PATON, T. A. and BAKER, A. J. 2006. Sequences from 14 mitochondrial genes provide a well-supported phylogeny of the Charadriiform birds congruent with the nuclear RAG-1 tree. *Mol. Phylogenet. Evol.*, **39** (3), pp. 657-667.

PEGG, G. G., SINCLAIR, B., BRISKEY, L. and ASPDEN, W. J. 2006. MtDNA barcode identification of fish larvae in the southern Great Barrier Reef, Australia. *Sci. Mar.*, **70**(S2): 7–12

PETHIYAGODA, R. and KOTTELAT, M. 2005. A review of the barbs of the *Puntius filamentosus* group (Teleostei: Cyprinidae) of southern India and Sri Lanka. *Raffles Bull. Zool. Suppl*, **12**:127-144

RAGAM, P. M. and RAMADEVI, K. 2005. An overview of the hill trouts (*Barilius* spp.) of Indian region. *Zoos' Print Journal*, **20**(4): 1847-1849

RASMUSSEN, R. S., MORRISSEY, M. T. and HEBERT, P. D. N. 2009. DNA barcoding of commercially important salmon and trout species (*Oncorhynchus* and *Salmo*) from North America. *J. Agric. Food. Chem.*, **57**: 8379-8385

RATNASINGHAM, S. and HEBERT, P. D. N. 2007. BOLD: The Barcode of Life Data System. *Mol. Ecol. Notes.*, **7**(3): 355-364

RUSCH, D. B., HALPERN, A. L., SUTTON, G., HEIDELBERG, K. B. and WILLIAMSON, S., et al. 2007. The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS. B.*, **5** (3): e77

SACCONE, C., DE GIORGI, C., GISSI, C., PESOLE, G. and REYES, A. 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene*, **238** (1): 195-209

SALGADO, J. HONORATO, B. and GARCIA- FONCILLAS, J. 2008. Review: mitochondrial defects in breast cancer. *Oncology*, **2**: 199–207

SAUNDERS, G. W. 2005. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philos. Trans. R. Soc. B.*, **360** (1462): 1879-1888

SHAW, J., LICKEY, E. B., BECK, J. T., FARMER, S. B., LIU, W., MILLER, J., SIRIPUN, K.C., WINDER, C. T., SCHILLING, E. E. and SMALL, R. L. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.*, **92**: 142–166

SCHLEI, O. L., LAFRENIERE, A. C., WHITELEY, A. R., BROWN, R. J., OLSEN, J. B., BERNATCHEZ, L. and WENBURG, J. K. 2008. DNA barcoding of eight North American coregonine species. *Mol. Ecol. Resour.*, doi: 10.1111/j.1755-0998.2008.02350.x

SMITH, M. A., WOOD, D. M., JANZEN, D. H., HALLWACHS, W. and HEBERT, P. D. N. 2007. DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. *Proc. Natl. Acad. Sci. U S A*, **104** (12): 4967-4972

STEINKE, D., ZEMLAK, T. S. and HEBERT, P. D. N. 2009. Barcoding Nemo: DNA-based identifications for the ornamental fish trade. *PLoS. ONE*, **4** (7): e6300. doi:10.1371/journal.Pone.0006300

TABERLET, P., COISSAC, E., POMPANON, F., GIELLY, L., MIQUEL, C., VALENTINI, A., VERMAT, T., CORTHIER, G., BROCHMANN, C. and WILLERSLEV, E. 2007. Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. *Nucl. Acids Res.*, **35**(3): e14

TABERLET, P., GRIFFIN, S. and GOOSSENS, B. et al. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res.*, **24** (16), 3189-3194.

TALWAR, P. K. and JHINGRAN, A. G. 1991. Inland fishes of India and adjacent countries. *Oxford & IBH, New Delhi*, **1**: liv+1–542, **2**: 543–1158 pp

TAMURA, K., PETERSON, D., PETERSON, N., STECHER, G., NEI, M., and KUMAR, S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, doi: 10.1093/molbev/msr121

TAUTZ, D., ARCTANDER, P., MINELLI, A., THOMAS, R. H. and VOGLER, A. P. 2003. A plea for DNA taxonomy. *Trends Ecol. Evol.*, **18**(2): 70–74

TELETCHEA, F., MAUDET, C. and HANNI, C. 2005. Food and forensic molecular identification: update and challenges. *Trends Biotechnol.*, **23** (7): 359-366

VALENCIENNES, A. 1844. Vol. 17. In: Cuvier, G. & A. Valenciennes, Histoire naturelle des poissons. Levrault, Strassbourg, xxiii+497 pp., pls. 487–51

VAN VELZEN, R., BAKKER, F. T. and LOON, J. J. A. 2007. DNA barcoding reveals hidden species diversity in *Cymothoe* (Nymphalidae). *Proc. Neth. Entomol. Meet*, **18**: 95-103

VENCES, M., THOMAS, M., BONETT, R. M. and VIEITES, D. R. 2005. Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Phil. Trans. R. Soc. B.*, **360**: 1859-1868

VINAS, J. and TUDELA, S. 2009. A validated methodology for genetic identification of tuna species (Genus *Thunnus*). *PLoS ONE*, **4** (10): e7606

WARD, R. D. and HOLMES, B. H. 2007. An analysis of nucleotide and amino acid variability in the barcode region of COI in fishes. *Mol. Ecol. Resour.*, doi: 10.1111/j.1471-8286.2007.01886.x

WARD, R. D., HANNER, R. and HEBERT, P. D. N. 2008. The campaign to DNA barcode of all fishes, FISH-BOL. *J. Fish Biol.*, **74**: 329-356

WARD, R. D., ZEMLAK, T.S., INNES, B. H., LAST, P. R. and HEBERT, P. D. N. 2005. DNA barcoding Australia's fish species. *Phil. Trans. R. Soc. B.*, **360** (1462): 1847-1857

WARES, J. P. and CUNNINGHAM, C. W. 2001. Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution*, **55**: 2455–2469

WILL, K. W., MISHLER, B. D. and WHEELER, Q. D. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.*, **54** (5): 844–851

WOESE, C. R. and FOX, G. E. 1977. Phylogenetic Structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U S A*, **74** (11): 5088-5090

XIA, J. XIA, K. and JIANG, S. 2008. Complete mitochondrial DNA sequence of the yellowfin seabream *Acanthopagrus latus* and a genomic comparison among closely related sparid species. *Mt. DNA*, **19** (4): 385–393

Zhang, D. X. and Hewitt, G. M. 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochem. Syst. Ecol.*, **25**: 99-120

VIII. ABSTRACT

Cataloguing of all life forms becomes an important requirement in biodiversity studies where in taxonomy plays an important role. Taxonomic classification of fishes is generally based on the morphometric characteristics and often these characters fail to distinguish closely related species as well as eggs, larvae and young ones of a species. Hence, there is a need to use a technique which helps to distinguish the species. DNA barcoding is one such technique which differentiates the fish species based on the DNA sequence. A DNA barcode is based on a short sequence of Cytochrome c oxidase sub unit 1 (COI) of mitochondrial genome which is species specific. In the present work, an effort is made to identify and validate some of the endemic cyprinids such as *Puntius filamentosus, Puntius assimilis, Barilius bakeri* and *Barilius canarensis* by barcoding technique, as these fish species were undergoing identity crisis and morphology based taxonomic approaches often fail to distinguish them. By analyzing the partial COI sequences the genetic distance calculated using Kimura 2 parameter within the species was lesser than that of between the species. The genetic distance between *B. canarensis* & *B. bakeri* was 14.6% and that of *P. filamentosus* and *P. assimilis* was 16.4%. The genetic distance between the genera *Puntius* and *Barilius* was 14.72%. Two of the genera were placed in different clades and *Channa striata* was unambiguously placed as outgroup in the phylogram. The clades after bootstrapping corresponded well with the expectations. We conclude that precise and accurate identification of *P. filamentosus, P. assimilis, B. bakeri* and *B. canarensis* could be done by using the barcode sequences of the mitochondrial DNA (COI) of these fishes.