CHAPTER 3

Molecular Characterisation using the mitochondrial CO1 and 16S rRNA genes

3.1 INTRODUCTION

The correct identification and characterization of living things is fundamental to biological science. The Earth biota is estimated to contain between 10 and 100 million eukaryotic species and the majority of species have yet to be identified (Stoeckle, 2003; Waugh, 2007). Traditional morphology based taxonomic procedures are not always sufficient for identification to the species level and can sometimes lead to misidentification (Cywinska et al., 2006). In this condition, a multidisciplinary approach to taxonomy especially by using the molecular data is become essential for species identification. The advent of DNA barcoding with rapid sequencing of mitochondrial DNA markers has given biologist a new tool for detecting and differentiating morphologically similar species (Krzywinski and Besansky, 2003; Bickford et al., 2006; Persis et al., 2009).

More than 50 years ago, starch gel electrophoresis of proteins was first used to identify species (Manwell and Baker, 1963). Nearly 40 years ago, single gene sequence analysis of ribosomal DNA were being used to investigate higher level evolutionary relationships (Woese and Fox, 1977) and mitochondrial DNA approaches dominated molecular systematics in the late 1970s and 1980s (Avise, 1994). Tautz et al. (2002, 2003) made the case for a DNA-based taxonomic system. DNA sequence analysis has been used for 30 years to assist species identifications, but different sequences have been used for different taxonomic groups and in different laboratories. Hebert et al. (2003) proposed that a single gene sequence would be sufficient to differentiate all or at least the vast majority of animal species and proposed the use of mitochondrial DNA gene cytochrome oxidase subunit I (cox1) as a global bio- identification system for animals and termed as DNA barcoding.
Fresh water fishes have received minimal systematic attention and many systematic problems remain. Accurate and unambiguous identification of fish and fish products, from eggs to adults, is important in many areas. It would enable species identification, assist in managing fisheries for long-term sustainability and improve ecosystem research and conservation. Mitochondrial DNA regions are mostly used as barcode genes for the characterisation of fishes (Briolay et al., 1998; Ward et al., 2005; Na-Nakorn et al., 2006; Filonzi et al., 2010). In the present study the mitochondrial DNA genes CO1 and 16S rRNA were used for the molecular characterisation of an endemic and endandered fresh water fish genus *Tor*.

3.1.1 Mitochondrial DNA (mtDNA)

A small portion of (<1%) of the DNA of eukaryotic cells is non-nuclear; it is located within organelles in the cytoplasm called mitochondria. The major features of mtDNA are: a) in general maternally inherited a haploid single molecule; b) the entire genome is transcribed as a unit; c) not subject to recombination and provides homologous markers; d) mainly selectively neutral and occurs in multiple copies in each cell; e) replication is continuous, unidirectional and symmetrical without any apparent editing or repair mechanism; and f) optimal size, with no introns present (Billington, 2003).

Even when two species are almost indistinguishable morphologically, they are likely to be easily distinguished genetically by using mtDNA. The vertebrate mitochondrial genome is composed of about 15 to 20 kb in different organisms, coding for 40 genes responsible for 2 ribosomal RNAs, 22 transfer RNAs, and 13 proteins essential in respiration (Ferris and Berg, 1987; Hartl and Clark, 1997; Okumus and Ciftci, 2003).

3.1.2 Phylogenetic Analysis

Several authors have reported studies on phylogenetic relationships, evolutionary pattern and genetic variability among species, subspecies and
populations in the genera of the family Cyprinidae using mitochondrial DNA sequences. Alves et al. (2001) observed significant mtDNA variation in the highly endangered cyprinid fish *Anaceypris hispanica* and identified four Management Units (MUs) among specimens representative of nine tributaries in the Iberian Peninsula. Nucleotide sequences of cyt b, tRNA genes, 12S rDNA and control region were used to assess the genetic and phylogeographic structure of *Acrossocheilus paradoxus* populations in 12 major streams from 3 geographical regions in Taiwan by Wang et al. (2000). The genealogical relationships and diversity within common carp (*Cyprinus carpio*) were well studied using allozyme, microsatellites and mitochondrial DNA markers (Kohlmann et al., 2003; Thai et al., 2004). Durand et al. (2002) utilized mitochondrial DNA cytochrome *b* region to study the phylogeny and biogeography of the family cyprinidae in the Middle East.

Phylogeographic studies on the cyprinid *Squalius aradensis* were carried out based on the sequence information of cytochrome *b* and microsatellites (Mesquita et al., 2005). Kyle and Wilson (2007) sequenced a 500 bp fragment of the mitochondrial cytochrome *b* gene from representative individuals from 26 harvested fish taxa from Ontario, Canada, focusing on species that support major commercial and recreational fisheries.

The analysis of mitochondrial DNA has proven a powerful tool for assessing intraspecific phylogenetic patterns in many animal species (Bernatchez et al., 1992; Avise, 1994). In initial studies with fishes (e.g. salmonids) universal primers of Cronin et al. (1993) for analysing the mtDNA, ND-1 and ND-5/6 segments were used as pointed by Avise et al. (1998), some 70% of studies carried out far involved analyses of mtDNA ‘phylogeography’.

The phylogeny and speciation modes in the European cyprinids (Briolay et al., 1998; Zardoya and Doadrio, 1999); cyprinid genus *Barbus* (Machordom and Doadrio, 2001); Eurasian and American cyprinids (Cunha et al., 2002);
East Asian cyprinids (Liu and Chen, 2003); finescale shiners of the genus *Lythrurus* (Cypriniformes: Cyprinidae), a group of 11 species of freshwater minnows widely distributed in eastern North America (Pramuk *et al*., 2007); East Asian cyprinid genus *Sarcocheilichthys* (Zhang *et al*., 2008) were studied by analyzing different regions of mitochondrial genome and were well documented. Mitochondrial DNA has attracted a lot of attention in many species, especially for population and evolutionary studies (Avise, 1994).

### 3.1.3 Concept of DNA barcoding

DNA barcoding is a technique which provides quick identification of species without involving the morphological cues. Morphologically indistinguishable species are likely to be easily distinguished genetically with the help of molecular markers. It uses a relatively small standardised DNA fragment as a tag to define or discover a species. An ideal DNA barcode should be normally a uniform short sequence of DNA (400-800 bp), able to be simply generated and used to characterize all the living organisms (Savolainen *et al*., 2005). The cytochrome *c* oxidase I (COI) region in the mitochondrial genome was demonstrated as an appropriate type-I molecular marker for discriminating closely related species (Hebert, 2004, 2005; Sengupta, 2013). Paul Hebert’s group was the first to design and use the short DNA sequences for biological identification at the University of Guelph, Canada. DNA barcoding uses the identification of one or a few regions in the genome to recognize all the species in a genus (Lahaye *et al*., 2008). DNA barcoding is essential for the molecular identification of already described species and the discovery of new species (Hebert *et al*., 2003).

### 3.1.4 The Potential Barcode Candidates

From 1990’s an increasing number of studies have been published making use of selected parts of mitochondrial and nuclear genome. Many studies have shown that short fragments of mitochondrial DNA, in particular the ‘barcode’
marker COI, are effective in diagnosing species (Hebert et al., 2004a, 2004b; Barrett and Hebert, 2005; Ward et al., 2005; Smith et al., 2008). Partial sequences of mitochondrial DNA genes especially 16S rRNA and Cytochrome C Oxidase I (COI) has proved more suitable than other gene sequences to resolve the phylogenetic relationships within the family in several group of eukaryotes including fishes owing to a large number of informative sites (Barucca et al., 2004).

3.1.5 16S rRNA (16S ribosomal RNA)

Several researchers from different parts of the world have used 16S rRNA gene sequences to study the phylogeny of different groups including fishes. Craig and Hastings (2007) have explained the molecular phylogeny of the groupers of the subfamily Epinephelinae (Serranidae) with a revised classification of the Epinephelini using 12S and 16S rDNA sequences. The phylogenetic studies of Sparks and Smith (2004) on cichlid fishes; Vinson et al. (2004) on Sciaenid fishes; Lopez et al. (2000) in Esocoid fishes; Wiley et al. (1998) on lampridiform fishes and Ilves and Taylor (2009) in Osmeridae are some other examples. Li et al. (2008) utilized the 16S rDNA sequences from 93 cyprinid fishes to reconstruct the phylogenetic relationships within the diverse and economically important subfamily Cyprininae.

3.1.6 COI (Cytochrome C Oxidase subunit I)

Cytochrome C Oxidase subunit I (COI) gene is supposed to be evolving faster than 16S rRNA which has been used widely in molecular taxonomy to resolve the phylogenetic relationships within the family and also to study intraspecific population genetic structure in several groups of eukaryotes including fishes. Klinbunga et al. (2005) utilized the COI sequences along with 16S and 18S rDNA to study the molecular taxonomy of Cupped Oysters (Crassostrea, Saccostrea and Striostrea) in Thailand. Well documented studies on population genetic structure were reported in several species including
crayfishes of the genus *Euastacus* in Australia (Ponniiah and Hughes, 2006); California red abalone (*Haliotis rufescens*) (Gruenthal *et al.*, 2007); Antarctic coastal krill *Euphausia crystallorophias* (Jarman *et al.*, 2002) and in atyid shrimp, *Paratya australiensis* (Baker *et al.*, 2004), utilizing COI sequences variations.

When fully developed, a COI identification system will provide a reliable, cost effective and accessible solution to the current problem of species identification as stated by Hebert *et al.* (2003a). Application of barcoding has detected the mislabelling of eight out of 22 Amazonian fish species in Brazilian fish market (Ardura *et al.*, 2010). Hebert *et al.* (2004a,b) have demonstrated that the COI region is appropriate for discriminating between closely related species across diverse animal phyla and this has been used for marine and freshwater fishes (Hajibabaei *et al.*, 2005; Steinke *et al.*, 2005; Ward *et al.*, 2005; Hubert *et al.*, 2008; Lakra *et al.*, 2009).

DNA barcoding is being employed to a large variety of organisms ranging from yeasts to vertebrates (Hebert *et al.*, 2004a, b; Hogg and Hebert, 2004; Moritz and Cicero, 2004). These results have prompted international efforts to standardize screening of species diversity and to accelerate the process of cryptic species identification. The real power of DNA barcoding lays on the ability of using the same locus to classify all species. In addition to the species identification, DNA barcoding has been used for identification of processed fish products (Smith *et al.*, 2008).

Accurate species identification is very important in case of morphological similar species for fisheries management, biodiversity studies and population dynamics. A global DNA-based barcode identification system that is applicable to all animal species will provide a simple, universal tool for the identification of fish species and products. When the reference sequence library is in place, new specimens and products can be identified by comparing their DNA barcode...
sequences against this barcode reference library (Lakra et al., 2011). The barcode system is based on sequence diversity in a single gene region (a section of the mitochondrial DNA cytochrome c oxidase I gene, COI).

DNA barcoding revealed a likely second species of Asian sea bass (barramundi) (Lates calcarifer). Ward et al. (2008a) strongly suggest that barramundi from Australia and from Myanmar are different species based on the sequencing of 650 base pair region of the mitochondrial COI gene. Spies et al. (2006) examined the variation at the mtDNA COI gene in 15 species of North Pacific skates and indicated that, a DNA-based barcoding approach may be useful for species identification. Zhang and Hanner (2011) used 229 DNA sequences of cytochrome oxidase subunit I gene (COI) from 158 marine fishes of Japan to test the efficacy of species identification by DNA barcoding and they concluded that DNA barcoding provides a new way for fish identification.

3.1.7 Indian scenario

Taxonomic ambiguity exists for several genera/ species of fishes and a proper identification is imperative for conservation management and trade. DNA-based approaches for taxon diagnosis exploiting DNA sequence diversity among species can be used to identify fishes and resolve taxonomic ambiguity including the discovery of new/cryptic species (Hebert et al., 2003). India has a rich natural heritage and nurtures a unique bio-diversity, placing it among the 12 most biodiverse countries. Out of 31,100 extant fish species, 2438 are known from Indian subcontinent (Froese and Pauly, 2009).

In India, microsatellite markers have been developed for 13 fish species, using primers from related species and have been found effective for population structure analysis (Gopalakrishnan et al., 2004; Mohindra et al., 2004, 2005). Research on the construction of a genomic library has also been carried out to develop microsatellite markers from Catla catla (Naish and Skibinski, 1998),
Endemic fish species from the Western Ghats such as *Labeo dussumieri*, *Horabagrus brachysom*a and *Puntius filamentosus* have been characterized using cytogenetic markers (Nagpure *et al.*, 2003). In marine fishes, species specific allozyme markers were developed to resolve the taxonomic ambiguity in marine catfish species such as *Tachysurus maculatus* and *T. subrostratus*.

DNA fingerprints of Indian major carps were obtained using banded Krait minor satellite 3 [BKM(3)] and M13 multilocus probes (Majumdar *et al.*, 1997) and using RAPD markers (Barman *et al.*, 2003; Das *et al.*, 2005b). Muneer *et al.* (2011) developed and used the RAPD and microsatellite markers for the characterization of genetic variability in the critically endangered yellow catfish *Horabagrus nigricollaris*. Menezes *et al.* (2002) identified diagnostic loci suitable for stock delineation and species identification among five nemipterid fish species from the Indian coast using allozyme analysis. Species-specific differences were identified between exotic *Clarias gariepinus*, non-native *C. macrocephalus* and native *C. batrachus* (Mohindra *et al.*, 2007) to detect introgression, if present.

Kushwaha *et al.*, (2002), characterized endangered mahseer species including *Tor khudree*, *T. mussullah*, *T. tor*, *T. putitora*, *Schizothorax richardsonii* and commercially important catfishes such as *C. batrachus* and *Heteropneustes fossilis*. They have reported the variation between *C. batrachus* and exotic *C. gariepinus* using cytogenetic markers in C- and nuclear organizer region (NOR) banding patterns for the first time (Nagpure *et al.*, 2000).

Some of the classical cytogenetic techniques like karyomorphology, chromosomal distribution of heterochromatin and different staining techniques have been utilized earlier for characterization of Indian fish species including *T. putitora* and *T. tor* (Khuda-Bukhsh and Nayak, 1982; Lakra, 1996; Kushwaha *et al.*, 2001). Mamta *et al.* (2009) performed the dual color fluorescence...
*in situ* hybridization (FISH) to study the simultaneous chromosomal localization of 18S and 5S ribosomal genes in the genus *Tor* for the first time. The genetic relatedness between five species of Mahseer group (*Tor putitora, Tor tor, Tor khudree, Tor mosal mahanadicus* and *Neolissochilus hexagonolepis*) was examined by Mohindra *et al.* (2007) for the first time using Random Amplified Polymorphic DNA (RAPD) analysis. Species-specific markers for two species of mahseer *T. khudree* and *T. mussullah* have been developed, which can be useful to resolve taxonomic ambiguity between these two species (Kushwaha *et al.*, 2001).

Lakra *et al.* (2010) studied the taxonomy and phylogeny of eight Channid fishes found in India using partial sequences of mitochondrial genes 16S rRNA and Cytochrome c Oxidase subunit I (COI). Persis *et al.*, (2009) analyzed the mitochondrial DNA, COI gene sequences for species identification and phylogenetic relationship among the 28 Indian Carangid fish species and the study proved the utility of mtDNA COI gene sequence based approach in identifying fish species at a faster pace. Study of Lakra *et al.* (2011) has strongly authenticated the efficacy of COI in identifying the fish species with designated barcodes. Molecular identification and phylogenetic relationships of seven Indian Sciaenids (Pisces: perciformes, Sciaenidae) fishes based on 16S rRNA and CO I mitochondrial genes was performed by Lakra *et al.* in 2009. Lakra *et al.* (2011) provided the first major barcode records for 115 commercially important Indian marine fish species belonging to 37 families.

The taxonomy and phylogenetic relationship of genus *Tor* is extremely confusing and a subject of debate due to the different types of morphological variations they exhibit based on the habitat (Silas *et al.*, 2005; Mohindra and Khare, 2007). Hora reviewed and evaluated the status of *Tor* species and concluded the taxonomic status of some species were more ambiguous. Menon (1992) studied the taxonomy of genus *Tor* - which led to many unanswered questions. Hence an efficient molecular method is very essential for species-level identification and conservation of genus *Tor*. 

94
3.2 OBJECTIVES

The objective of the present study is to perform the molecular characterisation and phylogenetic analysis of genus Tor using the mitochondrial COI and 16S rRNA genes.

3.3 MATERIALS AND METHODS

The fishes were collected from the sampling site using gill nets of different mesh size ranging from 8 mm to 22 mm, cast net and dip nets depending upon the depth and water velocity. The fishes were identified using the keys described by Talwar and Jhingran (1991), Menon et al., (1992) and Jayaram (2013). A small portion of tissue from the right side (fin clips of approximately 5 x 5 mm size) pectoral and pelvic fins was excised in a small tube and preserved in 99% Ethanol and labeled. Further the specimens were labeled and preserved in 10% formalin as voucher specimen for future reference.

3.3.1 Mitochondrial DNA analysis

3.3.1.1 Isolation of Genomic DNA

The total DNA was extracted from the tissue (fin clips) samples following Miller et al. (1988) with minor modifications. Tissue stored in alcohol was washed with TRIS buffer (pH 8.0) by spinning. The tissue sample was placed in 1.5 ml tube to which 500 μl of Solution 1 (50 mM Tris HCl (pH 8.0); 20 mM EDTA (pH 8.0); 2% SDS) was added. Homogenization of the sample was carried out with the sterile homogeniser and 5 μl of Proteinase K (20mg/ml). The solution was incubated at 55° C in a water bath for 2 hours (with occasional mixing) and chilled on ice for 10 minutes. 250 μl of Solution 2 (Saturated NaCl – 6M) was added and inverted several times for thorough mixing and again cooled on ice for 5 minutes. The sample was centrifuged at 8000 rpm for 15 minutes and the clear supernatant was carefully collected (~500 μl) with a wide-bore filter tip and transferred into a newly labelled 1.5 ml tube. If the supernatant was not clear the
centrifugation step was repeated again. 1.5 μl RNase (final conc. 20 μg/ml) was added in the supernatant and incubated at 37°C on a heating block for 15 minutes. Twice the volume (~1ml) of ice cold 100% molecular biology grade ethanol was added to precipitate the DNA and incubated overnight at -20°C. The next day, the solution was centrifuged at 11000rpm for 15 minutes and the supernatant was removed. The DNA pellet was rinsed in 250 μl of ice-cold 70% ethanol and centrifuged at 11000 rpm for 5 minutes and the supernatant was carefully removed and partially dried with the lid off at room temperature. The partially dried DNA was resuspended in 50-200 μl (depending on size of pellet) of TE buffer (pH-8) by gently pipetting the sample with a wide-bore filter tip until dissolved. The resulting DNA samples were stored at -20°C for further use.

Reagents required for Isolation of DNA

Stock solutions
1. **0.5M Tris Cl (pH-8.0)**
   - Tris base - 3.028g
   - Distilled water - 40ml
   - Adjust pH to 8.0 using HCl.
   - Make up the volume to 50ml, autoclave and store at 4°C.

2. **0.5M EDTA (pH-8.0)**
   - EDTA - 9.31g
   - Distilled water - 40ml
   - Adjust pH to 8.0 using NaOH.
   - Make up the volume to 50ml, autoclaved and stored at 4°C.

3. **10mM Tris Cl (pH-7.5)**
   - Tris base - 0.030g
   - Distilled water - 20ml
   - Adjust pH to 7.5 using HCl.
   - Make up the volume to 25ml, autoclaved and stored at 4°C.

4. **RNAase buffer**
   - 10mM Tris Cl (pH 7.5)- 10ml
   - 15mM NaCl – 30ml
   - Distilled water – 960ml
   - Autoclaved and stored at 4°C.

Working Solutions
1. **Solution 1:**
   - Tris-HCl (pH8.0) - 50mM
   - EDTA (pH8.0) - 20mM
   - SDS - 2%
   - Prepared in double distilled water.
   - Autoclave and store at 4°C

2. **Solution 2:**
   - NaCl solution (saturated) - (6M)
   - Prepared in double distilled water.
   - Autoclave and store at 4°C
3. **Proteinase K**  
Proteinase K - 20mg/ml  
Prepared in autoclave double distilled water and store at -20°C.

4. **TE buffer**  
Tris Cl (pH-8.0) - 10mM  
EDTA (pH-8.0) - 1mM  
Prepared in double distilled water.  
Autoclave and store at 4°C

5. **RNAase**  
RNAase - 10 mg/ml of RNAase buffer (autoclaved)

### 3.3.1.2 Agarose electrophoresis and visualization of bands

The extracted DNA was checked through 0.7% agarose gel (10x4 cm) electrophoresis with ethidium bromide incorporated in 1X TBE buffer. The gel casting unit was arranged according to the instructions of manufacturer. Agarose solution was boiled to dissolve agarose. Once cooled to approximately 50°C, 2.5 μl of Ethidium bromide solution was added to the solution and mixed thoroughly. The solution was poured into a gel casting plate with already adjusted gel comb and allowed to solidify at room temperature for 30 minutes. The comb was removed and the gel was then placed in the electrophoresis unit. 1X TBE buffer was poured into the electrophoresis unit as electrolyte. 2 μl of DNA solution was loaded together with 2 μl of sample loading buffer in each well. Electrophoresis was conducted at a constant voltage (80V) for 1 hr. After electrophoresis, the gel was observed in ultraviolet light and documented using the gel documentation system Image Master VDS (Pharmacia Biotech, USA).

**Reagents required for Agarose gel electrophoresis**

1. **TBE buffer 10X (pH-8.0)**  
   Tris base - 10.8g  
   Boric acid - 5.5g  
   EDTA - 0.75g  
   Make up the solution to 100ml with double distilled water.  
   Autoclaved and stored at 4°C

2. **Gel loading buffer**  
   Bromophenol blue - 0.5%  
   Glycerol (mol. grade) - 30%  
   Prepared in 1XTBE  
   Store at 4°C.

3. **Agarose solution (0.7%)**  
   Agarose - 0.21g  
   10X TBE - 3 ml  
   Distilled Water - 27 ml

4. **1X TBE buffer**  
   10X TBE - 10 ml  
   Distilled Water - 90 ml
5. **Ethidium bromide solution**  
   Ethidium bromide - 10mg  
   Distilled water - 2 ml

### 3.3.1.3 Estimation of quality and quantity DNA

The quality and quantity of the extracted DNA was checked in UV spectrophotometer (Beckman, USA) by taking the optical density (OD) at 260 nm and 280 nm. The quality was checked by measuring the ratio of absorbance at 260 nm and 280 nm (260/280). The value between 1.7 - 1.8 indicates good quality DNA without protein/RNA contamination. DNA quantification was done according to the following calculation: sample showing 1 OD at 260 nm was equivalent to 50 μg of DNA/ml. The OD of each DNA sample at 260 nm was measured and quantified accordingly.

### 3.3.2 PCR amplification of gene

The polymerase chain reaction (PCR) permits the synthetic amplification of a minute amount of DNA in millions of copies in a few hours (Mullis, 1990). PCR reactions were carried out in PTC 200 gradient thermal cycler (M.J. research, Inc., Watertown, Massachusetts, USA) employing specific universal primers for amplifications of partial sequences of mitochondrial DNA regions such as 16S rRNA and COI genes.

#### 3.3.2.1 Amplification of COI region

The partial sequence of cytochrome oxidase sub-unit I gene was PCR amplified using primers Fish F1 (5’ – TCA ACC AAC CAC AAA GAC ATT GGC AC - 3’) and Fish R1 (5’ – TAG ACT TCT GGG TGG CCA AAG AAT CA - 3’) (Ward et al., 2005). The amplifications were performed in 25 μl reactions containing 1x assay buffer (100 mM Tris, 500 mM KCl, 0.1% gelatin, pH 9.0) with 1.5 mM MgCl₂ (Genei, Bangalore, India), 5 pmoles of each primer, 200 μM of each dNTP (Genei, Bangalore, India), 1.5 U Taq DNA polymerase and 20 ng of
template DNA (Table 3.1). To check DNA contamination, a negative control was set up omitting template DNA from the reaction mixture. The following thermocycler conditions were used: initial preheat at 95°C for 3 min, denaturation 94°C for 30 s, annealing 50°C for 30 s, extension 72°C for 35 s, repeated for 29 cycles, followed by a final extension for 3 min at 72°C (Table 3.2).

### 3.3.2.2 Amplification of 16S rRNA region

The partial sequence of 16S ribosomal DNA was PCR amplified using primers L2510 (5’ – CGC CTG TTT ATC AAA AAC AT - 3’) and H3080 (5’ – CCG GTC TGA ACT CAG ATC ACG T - 3’) (Palumbi et al., 1991). The amplifications were performed in 25 μl reactions containing 1x assay buffer (100 mM Tris, 500 mM KCl, 0.1% gelatin, pH 9.0) with 1.5 mM MgCl₂ (Genei, Bangalore, India), 5 pmoles of each primer, 200 μM of each dNTP (Genei, Bangalore, India), 1.5 U Taq DNA polymerase and 20 ng of template DNA (Table 3.3). To check DNA contamination, a negative control was set up omitting template DNA from the reaction mixture. The following thermocycler conditions were used: initial preheat at 95°C for 3 min, denaturation 94°C for 30 s, annealing 58°C for 30 s, extension 72°C for 35 s, repeated for 29 cycles, followed by a final extension for 3 min at 72°C (Table 3.4).

### 3.3.2.3 Test Run and Purification of PCR product

About 3 μl PCR products along with marker (100 bp DNA ladder; Genei, Bangalore, India) were electrophoresed in 1.5% agarose gel (10x4 cm) using 1X TBE buffer for 90 minutes at constant voltage (80 V) and stained with ethidium bromide. The gel was visualized under UV transilluminator and documented using Image Master VDS (Pharmacia Biotech, USA). The remaining PCR product was purified using GeNeiTM Quick PCR purification kit (Genei, Bangalore, India) following the instructions given by the manufacturer.
Table 3.1. Composition of PCR reaction mix for CO1 gene

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Reaction volume (40μl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD Autoclaved water</td>
<td>32.2</td>
</tr>
<tr>
<td>PCR Buffer (10 X with 15 Mm MgCl₂)</td>
<td>4.0</td>
</tr>
<tr>
<td>dNTPs</td>
<td>0.2</td>
</tr>
<tr>
<td>Primer forward (pF)  (10 mM)</td>
<td>0.4</td>
</tr>
<tr>
<td>Primer reverse (pR) (10 mM)</td>
<td>0.4</td>
</tr>
<tr>
<td>MgCl₂ (25 mM)</td>
<td>0.8</td>
</tr>
<tr>
<td>Taq pol (3u/μl)</td>
<td>0.4</td>
</tr>
<tr>
<td>Template DNA (50 ng/μl)</td>
<td>1.6</td>
</tr>
<tr>
<td>Total Volume</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 3.2. Thermal regime for PCR amplification of CO1

<table>
<thead>
<tr>
<th>Steps</th>
<th>Conditions</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Temp.</td>
<td>Time</td>
</tr>
<tr>
<td>Initial denaturation</td>
<td>94° C</td>
<td>180 s</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94° C</td>
<td>30 s</td>
</tr>
<tr>
<td>Annealing</td>
<td>54° C</td>
<td>30 s</td>
</tr>
<tr>
<td>Extension</td>
<td>72° C</td>
<td>60 s</td>
</tr>
<tr>
<td>Final extension</td>
<td>72° C</td>
<td>600 s</td>
</tr>
<tr>
<td>Soak</td>
<td>4° C</td>
<td>-</td>
</tr>
</tbody>
</table>
### Table 3.3 Composition of PCR reaction mix for 16S rRNA

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Reaction volume (40μl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD Autoclaved water</td>
<td>32.2</td>
</tr>
<tr>
<td>PCR Buffer (10 X with 15 Mm MgCl₂)</td>
<td>4.0</td>
</tr>
<tr>
<td>dNTPs</td>
<td>0.2</td>
</tr>
<tr>
<td>Primer forward (pF) (10 mM)</td>
<td>0.4</td>
</tr>
<tr>
<td>Primer reverse (pR) (10 mM)</td>
<td>0.4</td>
</tr>
<tr>
<td>MgCl₂ (25 mM)</td>
<td>0.8</td>
</tr>
<tr>
<td>Taq pol (3u/μl)</td>
<td>0.4</td>
</tr>
<tr>
<td>Template DNA (50 ng/μl)</td>
<td>1.6</td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td><strong>40</strong></td>
</tr>
</tbody>
</table>

### Table 3.4. Thermal regime for PCR amplification of 16S rRNA

<table>
<thead>
<tr>
<th>Steps</th>
<th>Conditions</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Temp.</td>
<td>Time</td>
</tr>
<tr>
<td>Initial denaturation</td>
<td>94°C</td>
<td>180 s</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94°C</td>
<td>30 s</td>
</tr>
<tr>
<td>Annealing</td>
<td>47°C</td>
<td>40 s</td>
</tr>
<tr>
<td>Extension</td>
<td>72°C</td>
<td>60 s</td>
</tr>
<tr>
<td>Final extension</td>
<td>72°C</td>
<td>600 s</td>
</tr>
<tr>
<td>Soak</td>
<td>4°C</td>
<td>Forever</td>
</tr>
</tbody>
</table>
3.3.3 Gene sequencing

The cleaned up PCR products were used as the template for sequencing PCR. Nucleotide sequencing was performed by the dideoxy chain-termination method (Sanger et al., 1977) using ABI Prism Big Dye Terminator v3.1 Cycle Sequencing kit, (Applied Biosystems, USA). The composition of reaction mix is given in the Table 3.5. Terminators are dideoxynucleotides labelled with different coloured fluorescent dyes that will present different emission spectra on an electrophoresis gel illuminated by laser. In most cases, each PCR product was sequenced using the two amplification primers.

Table 3.5. Composition of reaction mix of gene sequencing

<table>
<thead>
<tr>
<th>Components</th>
<th>Vol. per reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDT (Big Dye Terminator - kit)</td>
<td>1.0 µl</td>
</tr>
<tr>
<td>Buffer (Supplied with Cycle Sequencing kit.)</td>
<td>1.5 µl</td>
</tr>
<tr>
<td>DNA (10-25ng/ µl)</td>
<td>1.0 µl</td>
</tr>
<tr>
<td>Primer (forward or reverse; 10pmol/µl)</td>
<td>0.5 µl</td>
</tr>
<tr>
<td>De ionized water</td>
<td>6.0 µl.</td>
</tr>
<tr>
<td><strong>Total volume</strong></td>
<td><strong>10.0 µl</strong></td>
</tr>
</tbody>
</table>

Cycle sequencing conditions were as follows:

**COI region:** 95°C for 30 s, 50°C for 5 s and 60°C for 4 min repeated for 25 cycles and finally stored at 4°C.

**16S rRNA region:** 95 °C for 30 s, 58 °C for 5 s and 60 °C for 4 min repeated for 25 cycles and finally stored at 4°C.

**Clean up for Sequencing**

The resulting cycle sequencing fragments were cleaned up adding 2 µl of 125 mM EDTA to each tube (PCR tube containing the PCR product) and mixed.
73 μl H₂O was added to this tube (final volume was adjusted to 100 μl) and it was then transferred to a 1.5 ml tube and 10 μl 3M sodium acetate (pH 4.6) was added. 250 μl of 100% ethanol was added and mixed gently. The tubes were incubated at room temperature for 15 min and it was spun at 12,000 rpm, at room temperature, for 20 min. The supernatant was decanted and 250 μl of 70% ethanol was added. Tubes were again spun for 10 min and the ethanol was decanted. The above step (ethanol wash) was repeated again. The supernatant was decanted and the pellet was air dried for 25-30 min.

### 3.3.3.1 Electrophoresis and visualization sequences

The cleaned up PCR products were sent for sequencing. The dried products were dissolved in 10% formamide and electrophoresed in a polyacrylamide gel and visualized using an AB 3730 XL capillary sequencer (Applied Biosystems). The products were visualized by laser detection of fluorescence emitted by different emission spectra of fluorescent labelled terminators. The raw DNA sequence information was generated as electropherograms read using DNA Sequencing Analysis Software Version 3.3 ABI (Applied Biosystems).

### 3.3.4 Sequence Analysis

The raw DNA sequences were edited using BioEdit sequence alignment editor ver. 7.0.5.2 (Hall, 1999). One example each of 16S and COI sequences were used to search Genbank for similar sequences using the Basic Local Alignment Search Tool (BLAST) available on the NCBI website (http://www.ncbi.nih.gov/BLAST/). The most similar sequences obtained from BLAST searches were added as outgroups for phylogenetic analysis to the data sets obtained for *Tor* in the present study. Multiple alignments of sequences were performed using CLUSTAL X ver. 2 (Larkin *et al.*, 2007) alignment editor. The alignment was checked and corrected manually. Aligned sequences were represented as rows within a matrix. Pairwise alignment of the sequence was
performed using dynamic programming. Phylogenetic trees were constructed from the multiple sequence alignments, based on the Neighbour Joining method and constructed using the alignment scores. The sequences (different haplotypes only) after their confirmation were submitted to GenBank (Appendix II) using a standalone multiplatform submission programme called “sequin” (www.ncbi.nlm.nih.gov/Sequin/index.html).

3.3.4.1 Phylogenetic analysis

Phylogenetic analysis was conducted using MEGA ver. 4 (Tamura et al., 2011). Sequence data was subsequently analysed using distance (Neighbour-Joining) and Maximum Parsimony methods. The sampling error of Neighbour-Joining and Maximum Parsimony Trees was analysed using bootstraps of 10,000 replicates where possible followed by the construction of Majority Rule Trees. Pairwise sequence divergence among populations was calculated according to Kimura two-parameter model (Kimura, 1980). The number and rate of transitions / transversions were also calculated using the program MEGA ver. 4.

3.3.4.2 Molecular Evolutionary Genetic Analysis (MEGA)

MEGA was used to investigate the evolutionary relationships between the species using homologous sequences. It is based on the statistical analysis of genes, the percentage of conservedness, variance and parsimony of the sequences. The distance between the sequence pair, average distance within and between groups were estimated. The estimation was accomplished using bootstrapping approach. Transaction and transversion type of substitution between the sequences were estimated. It was also used for inferring phylogenies by the distance based methods along with the bootstrap test.

3.4 RESULTS AND DISCUSSION

In the present study the mitochondrial CO1 and 16S rRNA regions were used for the characterization and phylogenetic analysis of Tor species collected
from seven geographically isolated locations throughout the Western Ghats. The genomic DNA was isolated and PCR amplification was carried out using universal primers. The amplified products are given in Figure 3.1 and 3.2 and the amplicons were sequenced. The obtained sequences were annotated and submitted in Genbank (Annexure II).

The partial sequence of the mitochondrial gene CO1 was sequenced and showed very distinctive sequences for each species. The size of the amplified product was approximately 655 bp. No insertions, deletions or stop codons were observed in any sequence. This supports the hypothesis that all the amplified sequences derive from a functional mitochondrial CO1 sequences. The absence of stop codons in the amplified sequences suggests that the nuclear DNA sequences originating from mitochondrial DNA sequences (NUMTs- Nuclear Mitochondrial DNA) were not sequenced (Ward et al., 2005; Lakra et al., 2011 and Persis et al., 2010). The occurrence of nuclear DNA sequences originating from mitochondrial DNA sequences has not been reported in Actinopterygii (Bensasson et al., 2001).

Sequencing of the 16S rRNA gene produced an average of 550 bp nucleotide base pairs per taxon. Simplicity and un-ambiguity were observed among the sequences of the 16S rRNA region, like mitochondrial CO1 region as there were no insertions, deletions and stop codons in the sequences.
Morphological similarity of the species and the difficulties involved in the identification of species make molecular information an invaluable investigative tool (Kyle and Wilson, 2007). Hebert et al., have shown that the analysis of short, standardized genomic regions (DNA barcodes) can discriminate morphologically recognized animal species (Hebert et al., 2003a) and the COI can serve as a uniform target gene for a bio identification system (Hebert et al., 2003b). Studies of Ward et al., (2008b) supported the scope of barcoding in identifying species and also the geographical variations expected within species. Out of fifteen fish species barcoded from Northern (Atlantic and Mediterranean) and Southern (Australasian) Hemisphere waters using COI sequences, Ward et al., (2008b) observed significant evidence of spatial genetic differentiation for this gene in two fishes; the silver scabbardfish (Lepidopus caudatus) and John dory (Zeus faber). Barcoding using the COI gene are also being used to verify the species identification of fish larvae,
especially of morphologically similar species (Pegg et al., 2006; Victor, 2007; Paine et al., 2008; Hubert et al., 2010; Kim et al., 2010; Baldwin et al., 2011 and Ko et al., 2013).

![Amplified product of 16SrRNA genes of Tor separated on 1.5% agarose gel](image)

**Figure 3.2. Amplified product of 16SrRNA genes of Tor separated on 1.5% agarose gel**

Lane 1- 6: *Tor malabaricus*    Lane 7- 10: *Tor khudree*    Lane 11- 14: *Tor mussullah*  
Lane 15- 20: *Tor sps.*       M- Marker gene (100bp DNA ladder)

The present analysis revealed the mean total nucleotide composition as A= 26.7%, T= 28.5%, G=17.5% and C= 27.3% in CO1 region. The mean total nucleotide composition was A= 31.7%, T= 21.6%, G=22.7% and C= 24.1% in 16S rRNA gene. The nucleotide frequencies of each species were calculated (Table 3.6 and 3.7). Like other teleosts (Gao et al., 2004; Ward et al., 2005; Persis et al., 2009), higher AT content than GC content was observed for both the CO1 and 16S mitochondrial fragments. More nucleotide changes were observed at the 3rd codon positions than the 1st codon. This reflected the fact that most synonymous mutations occurred at the 3rd codon, with a few at the 1st codon and none at the 2nd codon (Meyer, 1993; Zhang et al., 2009; Lakra et al., 2010).
Table 3.6. Nucleotide composition of COI region of Tor species

<table>
<thead>
<tr>
<th>Species</th>
<th>A (%)</th>
<th>T (%)</th>
<th>G (%)</th>
<th>C (%)</th>
<th>AT (%)</th>
<th>GC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tor khudree</td>
<td>26.1</td>
<td>28.6</td>
<td>17.8</td>
<td>27.5</td>
<td>54.7</td>
<td>45.3</td>
</tr>
<tr>
<td>Tor malabaricus</td>
<td>26.9</td>
<td>28.0</td>
<td>17.5</td>
<td>27.6</td>
<td>54.9</td>
<td>45.1</td>
</tr>
<tr>
<td>Tor mussullah</td>
<td>26.7</td>
<td>28.4</td>
<td>17.6</td>
<td>27.3</td>
<td>55.1</td>
<td>44.9</td>
</tr>
</tbody>
</table>

Table 3.7. Nucleotide composition of 16S rRNA region of Tor species

<table>
<thead>
<tr>
<th>Species</th>
<th>A (%)</th>
<th>T (%)</th>
<th>G (%)</th>
<th>C (%)</th>
<th>AT (%)</th>
<th>GC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tor khudree</td>
<td>31.6</td>
<td>21.7</td>
<td>22.8</td>
<td>23.9</td>
<td>53.3</td>
<td>46.7</td>
</tr>
<tr>
<td>Tor malabaricus</td>
<td>31.8</td>
<td>21.7</td>
<td>22.6</td>
<td>31.8</td>
<td>535</td>
<td>54.4</td>
</tr>
<tr>
<td>Tor mussullah</td>
<td>31.5</td>
<td>21.7</td>
<td>22.7</td>
<td>24.2</td>
<td>53.2</td>
<td>46.9</td>
</tr>
</tbody>
</table>

Analysis of COI revealed that out of 677 positions, 624 were conserved and 53 variable sites. 3 were singleton and 50 were parsimoniously informative. In 16S rRNA sequence, 533 were conserved, 48 variable sites, 16 were singleton and 32 were parsimoniously informative, out of 581 sequences. The overall mean distance among the species was 7.4% in the COI sequences. An average genetic distance within the species of Tor was 0.31% and the interspecific distance was 16.3%. The average genetic distance between species was estimated as 13.2% in 16S rRNA. The average distance within species was 0.39%.

A DNA based identification system depends on the ability to distinguish intraspecific from interspecific variation (Cywinska et al., 2006). As DNA
barcoding requires that intra-species DNA barcode variation should be substantially less than interspecific variation to allow accurate identification of individuals, the present result indicated the effectiveness of COI in identifying the species. Ward et al. (2005) sequenced two hundred and seven species of Australian marine fish, for a 655 bp region of the COI gene and found an average genetic distance of 0.003 within-species, 0.099 within genus, 0.154 within family, 0.221 within order and 0.232 within class. Lakra et al. (2008) reported very low genetic divergence within species of Scombrid fishes. The present work using the CO1 and 16S rRNA gene also supports Ward et al. (2005) by giving the less intra-species genetic distance than inter specific.

The effective application of DNA sequence data to molecular diagnostics depends on patterns of nucleotide substitution and the rate of variation among the sites (Blouin et al., 1998). In the present study the average transitional pairs were more frequent than the average transvertional pairs with an average ratio of 2.83 in CO1 sequence. As expected, average transitional pairs were more frequent than transversional pairs (with an average ratio of 1.9) in 16S rRNA sequence also.

Generally, among very close relatives such as a species of the same genus most of the changes are transitions at synonymous site. Much larger excess of transitions related to transversion is typically observed in the mtDNA of bony fishes (Ward et al., 2005). Transitions outnumbered the transversion in many research reports like DNA barcoding of Carangid fish, genetic differentiation of Macrodon ancyldon and mitochondrial DNA study of Sciaenidae (Santos et al., 2003; Vision et al., 2004; Persis et al., 2009). Among very close relatives such as species within a genus, most of the sequence changes are transitions at synonymous sites and among more distant relatives such as genera within a family or order transversions are more evident (Kocher et al., 1989). As the genetic distance increases, the transitions become saturated (Alves- Gomes et al., 1995; Kraus and Miya-moto, 1991).
An effective DNA-based identification system requires the satisfaction of three conditions: i) it must be possible to recover the target DNA from all species; ii) the sequence information must be easily analysed, and iii) the information content of the target sequence must be sufficient to enable species-level identification (Cywinska et al., 2006; Persis et al., 2009). All three of these requirements were met in the present study. Specimens of all species formed distinctive clusters. Finally, species boundaries were congruent with those established by morphological taxonomic work.

The efficacy of CO1 as a marker gene in identification and Phylogenetic relationship of the fish species have been proved by many authors (Smith et al., 2008; Lakra et al., 2009, 2011; Persis et al., 2009; Indu et al., 2012). The success in species diagnosis using CO1 gene reflects both the high rates of inter-species sequence change and the low intra-species rates in a wide range of animal taxa including ants, bats, birds, butterflies, crustaceans, fish and spiders (Hebert et al., 2004a, 2004b; Barrett and Hebert, 2005; Hajibabaei et al., 2005; Meyer et al., 2005; Ward et al., 2005; Hajibabaei et al., 2006; Cooper et al., 2007; Ratnasingham et al., 2007; Pyle et al., 2008; Zhang et al., 2011). The present study using the CO1 gene also showed the species level resolution with a high inter-species and a low intra-species sequence change. In the external appearance, most of the Tor species collected were looked alike and only from the close observation it was separated. Certain specimen was looking very different and the molecular analysis proved that the genetic content was the same.

All the three method of phylogenetic analysis (ML, MP and NJ) used in the present study gave the identical result. The NJ tree constructed using both CO1 (Figure 3.3) and 16S rRNA (Figure 3.4) gene revealed the identical phylogenetic relationship among the species. The phylogenetic relationship among the species was clearly established and similar species were clustered under same nodes while dissimilar species were clustered under separate nodes.
*Tor malabaricus*, *Tor mussullah* and *Tor khudree* were closer in both the trees with high boot strap support value.

Manimekalan (2000) and Easa and Shaji (2003) reported the presence of *Tor khudree* and *Tor mussullah* from the river Chaliyar. This analysis also proved the occurrence of *Tor mussullah* in Chaliyar. Easa and Basha (1995) also recorded *Tor khudree* during the exploration of fish diversity in the Karimpuzha tributary of Chaliyar. Baby et al., (2010) could not identify the *Tor khudree* from river Chaliyar and they mentioned that the *Tor khudree* recorded by Easa and Basha (1995) could be *Tor malabaricus* and not *T. khudree*. According to Arunachalam (IUCN, 2013) all *Tor khudree* recorded from Kerala, Karnataka and Tamil Nadu are *T. malabaricus* except for three populations in Chalakudy, Cauvery and Krishna basins.

*Tor khudree* and *Tor malabaricus* could be collected from different tributaries of Chalakudy, Chaliyar and Periyar river systems. The present study could identify and segregate the *Tor khudree* and *Tor malabaricus* based on the molecular data. This supports Silas et al., (2005) and confirmed *Tor khudree* and *Tor malabaricus* as separate species. In these three river systems, the co-existence of both *Tor khudree* and *Tor malabaricus* were observed. This is the first report on the presence *Tor malabaricus* from the tributaries of Periyar river, as there were no published literatures are available. In the present study, the molecular analysis using the mitochondrial COI and 16S rRNA genes was confirmed the presence of three species of *Tor- Tor khudree, Tor malabaricus* and *Tor mussullah-* in the Southern Western Ghat region.
Figure 3.3. Neighbour Joining (NJ) Phylogenetic tree ofTor species inferred from mitochondrial CO1 gene sequences
Figure 3.4. Neighbour Joining (NJ) Phylogenetic tree of *Tor* species inferred from 16S rRNA gene sequences
There was no significant geographical variation in the genetic makeup of Tor species observed. No intraspecific variation was seen among T. khudree and T. malabaricus based on the geographical location, but the same species were morphologically differed according to the varying geographical location.

It have been observed that some sequence of Tor samples taken from Cauveri river system and each species from Bhavani and Chalakudy were situated in the same clade with a high boot strap value of 100%. The multiple sequence analysis also showed 100% similarity between these sequence and they were differed from other sequences. To analyze the evolutionary isolation of these species and the level of divergence within species, K2P distance was calculated by averaging pairwise comparisons of sequence difference across all individuals by the Kimura 2-Parameter method (Kimura, 1980). There was no intraspecific distance observed between these species. But morphologically these specimens are looking different and this may be due to the effect of environmental factors of the various geographical locations. Ecology has an important role in predicting morphology. The Phylogenetic tree of all the Tor species present in NCBI also separates this group of fish (Figure 3.5). This group of fishes may be a new taxa and further study is needed for confirmation. DNA barcoding using the CO1 gene is also an efficient method even for the discovery of new species also (Ko et al., 2013).

Much controversy remains still regarding the role of molecular data within the taxonomic community particularly in DNA barcoding (Dayrat, 2005; Valdecasas et al., 2008; Packer et al., 2009; Holýński, 2010; Ebach, 2011; Mitchell, 2011; Shen et al., 2013). DNA barcoding also have its own limitations and this technology alone cannot define the taxonomy of a species. Hence, from the present study it can be concluded that an “integrative taxonomic” approach, which contain both morphological and molecular status, is necessary for the identification of Tor species.
Figure 3.5 Phylogenetic tree of all CO1 sequences of *Tor* from NCBI.
3.5 CONCLUSION

Mitochondrial CO1 and 16SrRNA genes were used for resolving the taxonomic ambiguities of genus Tor. The less intra-specific variation was observed than the interspecific variation in both CO1 and 16S rRNA genes. There is no significant intraspecific variation between the species from different geographical locations. The mitochondrial analysis also supports the existence of three distinct species of Tor in Southern Western Ghats, like the morphometric analysis.